

**Univerzita Karlova v Praze**  
**Přírodovědecká fakulta**

Studijní program: biologie

Studijní obor: zoologie



**Dmytro Omelchenko**

Genetická variabilita rodu *Alburnoides* v Ázerbajdžánu

Genetic variability of the genus *Alburnoides* in Azerbaijan

Diplomová práce

Školitel: RNDr. Miroslav Švátora, CSc.  
Konzultant Mgr. Radek Šanda, PhD.

Praha, 2016

**Prohlášení:**

Prohlašuji, že jsem práci vypracoval samostatně s použitím uvedené literatury. Práce je psána v jazyce anglickém a nebyla v ní provedena jazyková korektura. Anglický jazyk byl zvolen pro usnadnění vzniku budoucí odborné publikace a možnost případné konzultace se zahraničními odborníky. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

V Praze, 15. 08. 2016

.....

Dmytro Omelchenko

## Acknowledgements

Rád bych poděkoval mým rodičům za veškerou podporu během studia biologie. Děkuji svému školiteli **Miroslavu Švátorovi** za navržené téma, konzultace a průvod akademickým prostředím Univerzity Karlovy.

Chtěl bych také poděkovat svému konzultantovi **Radkovi Šandovi** za cenné rady a připomínky, finanční podporu a též za vedení v průběhu praktické části této studie.

Taktéž děkuji **Zuzaně Musilové** za pomoc se sestavováním této práce a přípravu fylogenetických analýz a v neposlední řadě za přátelský přístup.

Dále děkuji **Akifu Gulievu** a kolektivu státní univerzity v Baku za pomoc v organizaci terénní práce a sběru materiálu.

Chci poděkovat **Danielu Fryntovi** za konstruktivní připomínky k interpretaci výsledků.

Rad bych poděkoval **Zuzaně Varadinové** a **Dovile Barčytě** za pomoc a školení v laboratoři.

Také děkuji **Zdeňku Lerchovi**, **Milanu Kaftanovi**, **Evě Landové** a **Barbarě Kaftanové** za spoluúčast při expedici do Ázerbajdžánu.

**Abstract:**

The Caucasus region is characterized by high rate of endemism and taxa richness of fishes. Azerbaijan is a country situated on the border between Europe and Asia with rivers flowing in the Caspian Sea. Natural environment of this country is represented by various habitats with diverse ichthyofauna. This region is very attractive for biogeographical studies because it lays on the border of two different ecoregions. Even on the modern stage of scientific cognition, there is still lack of data about freshwater fishes from that region. Spirlins or rifle-minnows (*Alburnoides* Jeitteles, 1861, Actinopterygii, Cyprinidae) is a genus of small freshwater fishes and it has been chosen as a focus of this thesis because of numerous reports of new species from surrounding countries. The presented thesis is one of the seldom molecular studies trying to reveal the taxonomical situation within the genus *Alburnoides*, describe the phylogenetic relationships between geographically isolated populations, and provide biogeographical implications for fishes in the Caspian Sea river basins. Both mitochondrial (cytochrome b, cytochrome oxidase subunit I) and nuclear (RAG1, rhodopsin) markers were used in the study and the Maximum Likelihood, Maximum Parsimony and Bayesian phylogenetic analyses were performed. Further, the method of Molecular Clock was applied to estimate divergence between lineages inside genus *Alburnoides* and for subsequent constructing of biogeographical scenario. High level of genetic variability has been found within the genus *Alburnoides* in Azerbaijan. Examined individuals represent three well-supported major clades in phylogenetic trees, some of them (but not all) corresponding to the geographically isolated localities. The population structure and haplotype distribution within the *Alburnoides* genus cannot be explained only by geographical isolation. There is a putatively undescribed species, *Alburnoides* sp. 5 in Ciscaucasian region, a lineage which probably diverged as a result of a vicariant event. Identified lineage richness within the samples from the Talysh Mountains hydrological network probably show the evidence for the existence of glacial refugium in the region. The regions in the North (Greater Caucasus hydrological network) were genetically uniform, while the localities in the South (Talysh Mountains hydrological network) were composed of two different lineages. This brings an evidence for putative colonization events from the North to the South, which might have been facilitated by the known currents in the Caspian Sea. Newly reported invasive species for the Azerbaijan ichthyofauna, *Hemiculter leucisculus*, has been registered.

**Keywords:** *Alburnoides*, biogeography, fish, genetic variability, phylogeny.

## Abstrakt:

Region Kavkazu je charakterizován vysokou úrovní endemizmu a bohatstvím taxonů ryb. Ázerbájdžán je stát, který je situovaný na rozhraní mezi Evropou a Asií, a jeho řeky tečou do Kaspického moře. Přírodní prostředí tohoto státu představují různorodé habitaty s diversifikovanou faunou a daný region je velmi atraktivní také pro biogeografický výzkum, jelikož leží na rozhraní dvou různých ekoregionů. I v současné době stále chybějí podrobné znalosti o sladkovodních rybách z toho regionu. Ouklejký (*Alburnoides* Jeitteles, 1861, Actinopterygii, Cyprinidae) jsou rodem drobných sladkovodních ryb, které byly vybrány jako objekt výzkumu také kvůli početným objevům a popisům nových druhů z okolních států. Předkládaná práce je jedná z mála molekulárních studií, která se pokouší vyřešit taxonomickou situaci uvnitř rodu *Alburnoides*, popsat fylogenetické vztahy mezi jednotlivými populacemi a poskytnout biogeografické interpretace pro ichtyofaunu z povodí řek Kaspického moře. V práci byly použity jak mitochondriální (cytochrom b, cytochrom oxidáza podjednotka I) tak i jaderné (RAG1, rodopsin) markery a analýzy maximální věrohodnosti (Maximum Likelihood), maximální parsimonie (Maximum Parsimony) a Bayesovské analýzy byly použité k rekonstrukci fylogenetických a fylogeografických vztahů. Dále byla použita metoda molekulárních hodin (Molecular Clock) pro odhad času divergence mezi jednotlivými liniemi uvnitř rodu *Alburnoides* a pro následnou rekonstrukci biogeografického scénáře. Byla potvrzena výrazná genetická variabilita uvnitř rodu *Alburnoides* v Ázerbájdžánu. Zkoumaní jedinci ve fylogenetických stromech tvoří tři dobře podpořené linie (clades); někteří z těchto linií jsou vázány výlučně na geograficky vzdálené a izolované lokality. Populační struktura a distribuce haplotypů nemůže být ovšem vysvětlena jen na základě geografické izolace. Rodová linie nepopsaného druhu *Alburnoides* sp. 5 z Předkavkazska se hypoteticky oddělila následkem vikariantní události. Zjištěné bohatství rodových linií mezi vzorky z hydrologické sítě Talyšských Hor je pravděpodobně důkazem existence glaciálního refugia v regionu. Populace na severu Ázerbájdžánu (hydrologická síť Velkého Kavkazu) jsou geneticky uniformní, zatímco lokality na jihu (hydrologická síť Talyšských hor) byly obývány dvěma nezávislými liniemi ouklejek, což by mohlo být následkem kolonizačních událostí směrem ze severu na jih, které mohly být také usnadněny díky proudům v Kaspickém moři. Invazivní druh *Hemiculter leucisculus* nedávno popsáný pro faunua Ázerbájdžánu byl zaregistrován také během terénních sběrů použitých pro tuto práci.

Klíčová slova: *Alburnoides*, biogeografie, fylogeneze, genetická variabilita, ryby.

## Contains:

1 Introduction.....	8
2 Literature overview .....	9
2.1 Overview of the genus <i>Alburnoides</i> . Insights into morphology, taxonomy and evolution. .....	9
2.2 Biogeography of <i>Alburnoides</i> genus.....	11
2.3 Fossil record of Cyprinid fishes.....	13
2.4 Biogeographical delineation of the research region .....	14
2.5 Geological history of Caucasus and Caspian Sea .....	19
3 Main goals of present study .....	25
4 Materials and methods .....	26
4.1 Samples collection .....	26
4.2 DNA extraction and PCR .....	27
4.3 Dataset preparation .....	28
4.3.3 Datasets for MP, ML and Bayesian methods. ....	28
4.3.2 Haplotype network dataset.....	29
4.3.3 Concatenation of genes .....	29
4.4 Model testing .....	29
4.5 Maximum likelihood method.....	30
4.6 Maximum Parsimony.....	30
4.7 Bayesian analysis.....	30
4.8 Haplotype network.....	31
4.9 Molecular clock method .....	31
5 Results.....	33
5.1 Genetic variability of the genus <i>Alburnoides</i> inferred from mitochondrial cytochrome b analysis.....	34
5.2 Genetic variability of genus <i>Alburnoides</i> inferred from the analysis of mitochondrial cytochrome oxidase subunit I gene.....	38

5.3 Genetic variability of genus <i>Alburnoides</i> inferred from the analysis of nuclear recombination activating gene I.....	42
5.4 Genetic variability of genus <i>Alburnoides</i> inferred from the analysis of nuclear rhodopsin gene.....	46
5.5 The analysis of concatenated mitochondrial and nuclear genes .....	48
5.6 Molecular clock method .....	50
5.7 Haplotype network.....	52
6 Discussion.....	54
6.1 Discussion of cytochrome b signals.....	55
6.2 Discussion of cytochrome oxidase subunit I signals .....	57
6.3 Controversy of the Eich N+S clade .....	59
6.4 Biogeographical scenario.....	60
6.5 Putatively undescribed species, <i>Alburnoides</i> sp. 5 in Ciscaucasian region .....	62
6.6 <i>Hemiculter leucisculus</i> new invasive species for Azerbaijan fauna .....	64
7 Conclusions.....	65
8 Literature.....	66
9 Appendix.....	73

## 1 Introduction

Caucasus in general, and Azerbaijan rivers in particular, represent a region with high rate of endemism and species richness. This statement is only valid for Europe, as number of endemic freshwater fishes in Caucasus region is notably lower (20 – 27 endemic freshwater fish species (Abell et al., 2008)) comparing to the number of endemics in tropical regions, such as Amazon or Great African Lakes (between 196-387 endemic fish species; Abell et al., 2008). The total number of freshwater fish species is 27-35 species for the almost whole Azerbaijan and 36-43 species for southern part (Freyhof et al., 2014). This is still the highest value within the Eastern Mediterranean region (Freyhof et al., 2014), nevertheless it is lower when compared to the entire Peri-Mediterranean area with 110 species (Reyjol et al., 2007).

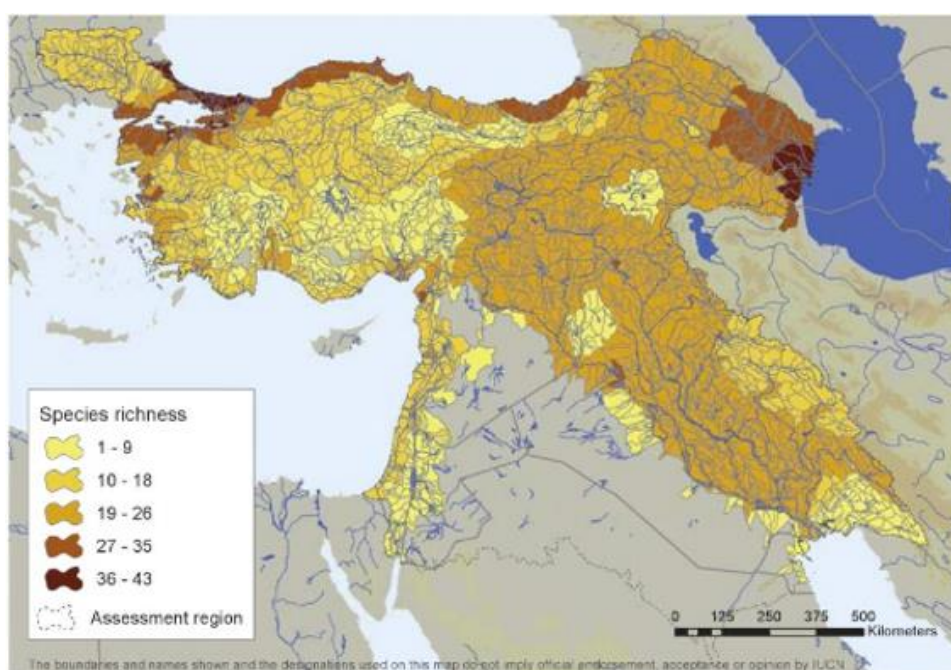


Figure 1. Species richness in the Eastern Mediterranean region (Freyhof et al., 2014).

However, the biodiversity of Azerbaijan might still be underexplored and recently, 18 new species have been described only in the genus *Alburnoides* from the neighboring countries, like Iran and Turkey after Abell's calculations (2008).

Nevertheless, I would recommend to consider the number of species in Caucasus region as approximate estimations only, due to the lack of sufficient valid scientific data from that region. In this thesis I aim to clarify the situation with the respect to one fish genus, spirilins and rifle minnow (*Alburnoides* Jeitteles, 1861).

Theoretically lineage richness in particular biogeographical region is closely connected



with existence of glacial refugia (Reyjol et al., 2007). Between those refugia: Balkan, Caucasus, Iberian and Italian peninsulas are widely accepted (Hewitt, 1999). I hypothetically consider those geographical regions as hotspots of speciation processes.

## **2 Literature overview**

### **2.1 Overview of the genus *Alburnoides*. Insights into morphology, taxonomy and evolution**

Rifle minnows or spirilins (*Alburnoides* Jeitteles, 1861, Actinopterygii, Cyprinidae) are relatively small freshwater cyprinid fishes widely distributed in the clear, fast flowing rivers in European part of Eurasia. Spirilins further inhabit some Asia Minor and Central Asia drainages that belong to Black, Caspian, Aral Sea and Persian Gulf basins. Spirilins can be easily distinguished from other cyprinid fishes by black spots on both sides of the lateral line and orange base of the pectoral, pelvic and anal fins. Currently, there are 33 valid species in the genus *Alburnoides* (Eschmeyer, 2016). For more information, see Appendix – List of valid *Alburnoides* species. Number of species in the genus has increased dramatically in last ten years: 18 further species had been recently described (in addition to previously known 15 species). Descriptions of these new species are based solely on the morphological characteristics, such as fin-rays and vertebrae counts (Bogutskaya & Coad, 2009; Coad & Bogutskaya, 2009; Coad & Bogutskaya, 2012; Turan et al., 2013, 2014, 2016; Mousavi-Sabet et al., 2015). However, several molecular studies have recently confirmed great diversity within the genus (Seifali et al., 2012; Stierandová et al., 2015; Jouladeh-Roudbar et al., 2016).

Similar to other cyprinids, rifle minnows are morphologically characterized by a tooth-bearing pharyngeal bones, basioccipital with enlarged ventroposterior process, a protrusile mouth actuated by a kinethmoid, toothless palate and jaws, highly developed fifth levator muscle which originates from exceptionally deep subtemporal fossa (Howes chapter I in Wienfield and Nelson, 1991).

Taxonomical position of the genus *Alburnoides* inside the family Cyprinidae seems to be clear and well supported by numerous morphological (Chen et al., 1984; Howes, 1991) and molecular studies (Briolay et al., 1997; Cunha et al., 2001; Gilles et al., 1998; Gilles et al., 2001; Durand et al., 2002; He et al., 2007; Wang et al., 2012). Chen et al. (1984) had placed the genus *Alburnoides* into the Leuciscinae subfamily. That placement of spirilins inside the Leuciscinae subfamily has no disputable points, but Howes (1991) provided a cladistic analysis and defined the synapomorphies of *Alburnoides*, *Alburnus*, *Hemiculter* etc.,

extracting these genera from the subfamily Leuciscinae into the Alburninae subfamily and considering them as a monophyletic group. To support this idea Howes mentioned that representatives of Alburninae share following morphological characteristics: laterally compressed parasphenoid with outwardly curved ascending processes, elongate basicranial foramen placed between parasphenoid and basioccipital, lateral ethmoid with truncated wings, extensive carotid foramina separated in the midline by only a narrow portion of the parasphenoid, frontals transversely convex, ventral border of coracoid serrated or fretted, cranially originating intermuscular bones, pelvic bone expanded and orientated vertically, expansion of neural spine of PU2 (pre-ural centrum) and pectoral axial lobes in some taxa (Howes, 1991).

Contrarily, several further morphological studies are not congruent with the hypothesis of Howes. Cavender and Coburn (1992) recognized only two subfamilies (Leuciscinae and Cyprininae) and that classification is based on the simple conclusions that members of the Cyprininae subfamily have a «head usually rigid when feeding and slow swimming movements in feeding», while members of the Leuciscinae have a “head lifting mechanism when feeding and often feeding with rapid swimming movements» and on the presence or absence of barbells (Cavender and Coburn, 1992).

In the last decade the molecular approach becomes essential to test systematic, phylogenetic and evolutionary hypotheses. There are numerous studies trying to provide evidence for taxonomy of the Cyprinidae family based on mitochondrial and nuclear markers. Zadroya and Dodario (1998) in research based on cytochrome b usage concluded that Leuciscinae and Alburninae could be fused into the one subfamily and that there are no strong evidences supporting the monophyly of the Alburninae subfamily. Other authors (Gilles, 1998, 2001; Cunha, 2001) made similar conclusions using mitochondrial markers namely cytochrome b and 16S.

More recent molecular phylogenetic hypothesis based on nuclear markers (He et al., 2007) also doesn't support separation of the Alburninae lineage from the Leuciscinae subfamily and rather places the genus *Alburnoides* and other close genera as members of big polyphyletic subfamily Leuciscinae. Nuclear markers became popular in the molecular phylogenetic due to some obvious limitations of mitochondrial markers for solving phylogenetic problematics based on high number of taxa, such as: Mitochondrial genome is maternally inherited, has higher mutation rate and theoretically has a shorter time towards the coalescence. It may lead to formatting doubtful groups of species which with an introgression and hybridization processes are not reflecting the true phylogenetical relationships within the

groups of interest (He et al., 2007). But mitochondrial genome has a highest mutation rate what makes it better for solving intraspecific phylogenetic relationships.

In conclusion, in the monolithic textbook by Nelson (2006) devoted to the fishes of the world, problematics of Cyprinid taxonomy had been discussed and author concluded that it is still relevant and disputable. He included the *Alburnoides* genus inside the Leuciscinae subfamily and I will consider this position as valid in this study.

## 2.2 Biogeography of *Alburnoides* genus

Biogeography of freshwaters in the Palearctic region has been reviewed and discussed in numerous studies (Berg, 1949; Bănărescu, 1975; Bănărescu & Coad, 1991; Abell 2008, Naseka 2010). In accordance to Bănărescu and Coad (1991) rivers and lakes of Eurasia may be divided in seven major areas according to their geography, history and ichthyofauna.

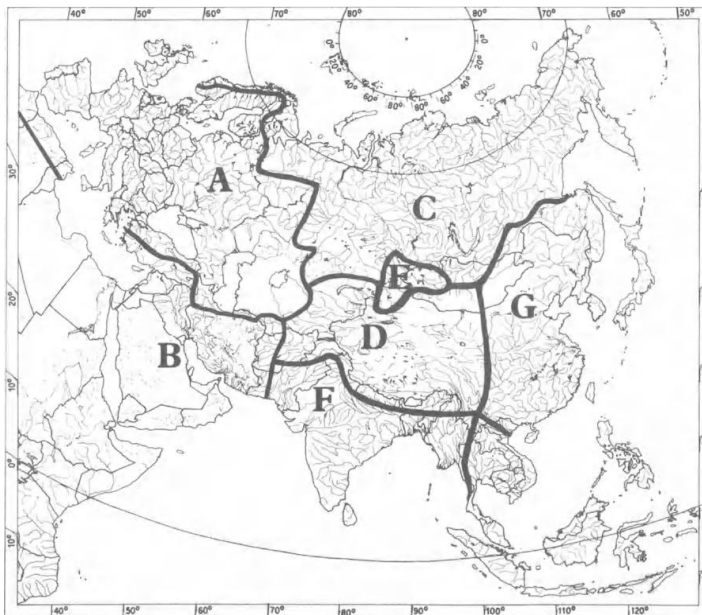


Figure 2. Major freshwater areas in Eurasia (South-East Asia not included). Where: (A) Europe, including North-West Africa, central and northern Anatolia, Transcaucasia, the Caspian Sea basin of Iran, and the lower reaches of the Aral Sea basin; (B) South-west Asia, including southern Iran, Tigris-Euphrates basin, southern Anatolia, the Levant and the Arabian Peninsula; (C) Siberia, with Arctic Ocean drainages and Northern Pacific Ocean drainages; (D) Central or High Asia, including internal basins and the upper reaches of rivers flowing through the other areas , e.g. the Huang He and Yangtze, the Mekong, Irrawaddy, Salween, Brahmaputra, Ganges and Indus, Amu Darya and Syr Darya; (E) Western Mongolia, with refugial areas of the Sayan and Altai mountains; (F) South Asia, comprising the Indian subcontinent and Burma (basins of the Salween, Sittaung and Irrawaddy and the Western slope of the Malayan Peninsula); (G) East Asia, including the Amur basin south to Song Koi, including Sakhalin, Japan, Taiwan and Hainan Dao. (from Wienfeld and Nelson, 1991, chapter 4 by Bănărescu and Coad).

The geographic distribution of *Alburnoides* genus corresponds largely to the distribution of the most widely distributed species, *Alburnoides bipunctatus*. Spirilins inhabit Europe and South-West Asia from presented biogeographical delineation. The native distribution of spirilins covers the most southern parts of Scandinavia to the Urals, south-west to Pyrenees, eastwards through the Alps, Balkans and Danube drainage. In addition to that, eastward almost entire Black Sea basin, (the rivers of Asia Minor and the presence of *A. bipunctatus* there is still under the question mark) and upper parts of rivers flowing from the north into Caspian and Aral Sea. We can consider *A. bipunctatus* as a common or European spirilin, because of its wide distribution area. Also it comes through all over Europe with unitary suite of morphological and genetic characteristics, at least, so far there are no reports of new species inside that area. Similarly, the species *Alburnoides eichwaldii* has a second largest distribution among rifle minnows. It has been found in the rivers of Asia Minor, southern parts of Caspian and Aral Sea basins, and inflows of Tigris-Euphrates river system.

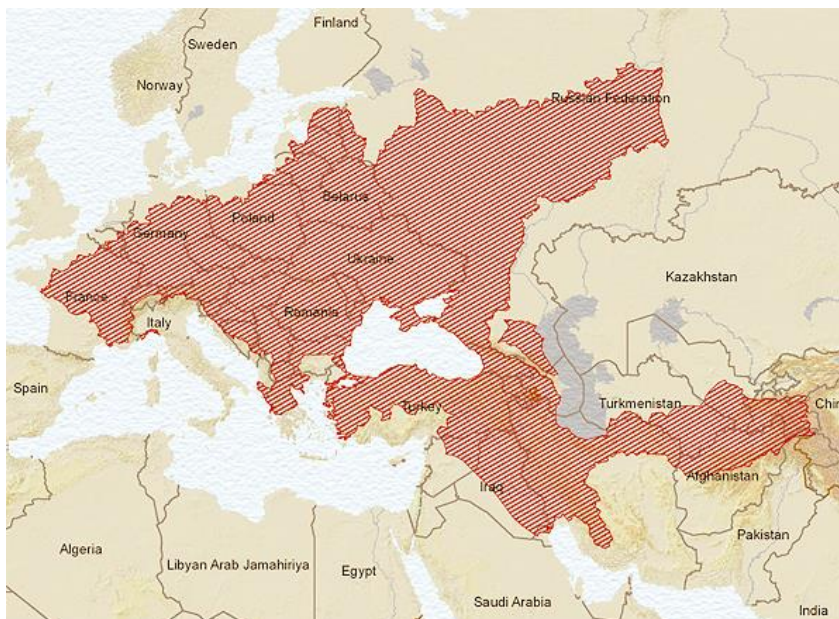


Figure 3. *Alburnoides bipunctatus* distribution map (adapted from <http://www.ittiofauna.org/>). Distribution of another representatives of genus is congruent to *A. bipunctatus*.

Kottelat and Freyhof (2007) reported only three species from European freshwaters. Except for the most common European spirilin (*A. bipunctatus*), there is the Ohrid spirilin (*A. ohridianus*) and the Prespa spirilin (*A. prespensis*) from the Balkan region. This region is known for the high rate of endemism and number of species. The authors mention that the number of species in the genus *Alburnoides* is probably underestimated, mostly due to numerous reports from the eastern edges of the *Alburnoides* distribution. It seems likely that the Caucasus and the Middle-East regions could serve as a speciation centers due to the

specifics of their geological history (Durand, 2002) as these regions have been previously identified as a “cross-road” for cyprinid fish dispersal (Durand, 2002).

Many representatives of the closely related taxa (*Alburninae* sensu Howes), namely the genera *Hemiculter* and *Metzia* inhabit also Eastern Asia, which is outside the usual distribution as compared to *Alburnoides*. For that reason, we are currently considering distribution of the *Leuciscinae* subfamily as disjunctive. European and Eastern Asian parts of the distribution area are disrupted by Central Asia region, where no *Alburninae* can be found. This distributional pattern is typical also for others cyprinids and it can be found in both cyprinine and leuciscine lineages (e.g. *Phoxinus*, *Barbus* (Howes, 1991). The explanation of the origin of such disrupted distribution patterns depends mostly on the general biogeographical approaches and remain unclear to certain extent (Howes, 1991). There are still discussions between dispersalists and vicariographers.

### **2.3 Fossil record of Cyprinid fishes**

In order to have better understanding of this historical events underlying modern distribution patterns of *Alburnoides* and other cyprinids, closer focus on the fossil record and modern population structure is required. By integration of these approaches, we can illustrate the evolutionary processes on the larger scale and in the broader context.

Cavender (1991) studied the fossil record of the cyprinid fishes in Europe and North America and concluded that radiation of cyprinids at temperate latitudes had taken place in the Late Paleogene. With additional reference to Prothero (1989) he assumed that Eocene terminal extinction event which eliminated many aquatic and terrestrial animals had been inducing climate stress for 10 million years with decreasing of the global temperature by 10 degrees. This event took place between 40 and 30 mya and these dramatic climate changes triggered changes of biota. For example, tropical forests of the mid-Eocene were replaced by swamps and reed marshes in the Late-Eocene, and the tropical mammals inhabited forests were substituted by grazing herbivores of the open landscapes (Cavender, 1991).

The Oligocene cyprinids were represented mostly by the leuciscine ancestors and their pharyngeal teeth structure suggest the insectivorous diet (Cavender, 1991). Hypothesis of cyprinids having the centre of origin in China and South-East Asia is widely accepted (Bănărescu & Coad, 1991). In the recent ichthyofaunal record, cyprinid fishes in this particular biogeographical region exist in huge number of genera and species. It seems to be likely that ancestors of modern cyprinid genera dispersed from that region westward to Europe (Bănărescu & Coad, 1991). During such invasion, some populations were captured by

geomorphological processes caused by tectonic movements.

Gilles (1998) estimated the time of the first cladogenesis within Cyprinidae as 43 to 24.6 mya based on the mitochondrial genetic signals (cytochrome b and 16s rDNA). If we will apply this to the events in the Eocene, we can assume that they might played a very important role in cladogenesis. In this paper, authors assumed that their evaluation likely confirms the hypothesis of Almaça assuming that extant European cyprinids have diverged 35 million years ago (Almaça, 1976).

Bianco (1990) proposed that Late Messenian salinity crisis (between 6.5 to 5.3 mya) led to the close contact between populations of the coastal freshwater fishes and, as a result, their spread and numerous speciation events. This event is considered as one of the most influential events with large impact on the European freshwater ichthyofauna in general. In this time, leuciscine biodiversity became restricted to limited freshwater areas, and as a result the new morphs and introgressions appeared (Gilles et al., 1998).

According to Gilles (1998), *Alburnoides* diverges from other leuciscines by 6.98% of the average sequence divergence of cytochrome b. They have further estimated the probable time the basal divergence within leuciscine lineage to happen between 17.5 and 9.9 mya. These estimates are, however, very poorly applicable for evolutionary story of genus because of numerous introgressions that presumably took place in the evolutionary history of leuciscine lineage. In some cases, their appearance might be explained by marine regression during glacial maximum of Pleistocene. However, authors stated: “taxa that are known to be involved in hypothetical introgressions have not been taken into account in these calculations” (Gilles et al., 1998). Main goal of this thesis is to extend knowledge about this problematic.

## **2.4 Biogeographical delineation of the research region**

Azerbaijan is a country which is situated on the border of Eastern Europe and Southwestern Asia between Black and Caspian Sea. The uniqueness of that particular part of the globe relays in the fact, that it is occupying an entire edge of the geographical Europe. From National Geographic Atlas of the world: "A commonly accepted division between Asia and Europe is formed by the Ural Mountains, Ural River, Caspian Sea, Caucasus Mountains, and the Black Sea with its outlets, the Bosphorus and Dardanelles." (Zar, 1982). Azerbaijan has three big mountain ranges: Greater Caucasus on the north of the country (border with Russia), Lesser Caucasus with Talysh Mountains on the south (border with Armenia and Iran) and Transcaucasian Depression with Kura-Aras Lowlands located in-between those ridges. The

Kura river (Mtkvari) collects water from drainages flowing down from mountains. Kura is the longest river in the Caucasus region. Middle and lower parts of Kura river are outside my current focus because they are situated in the semi-arid central region of Azerbaijan. Riffle minnows, as it was mentioned before, are rheophilic fish with strong affinity to clear water and high oxygen concentrations, so they almost absent in the central regions.

Caucasus region belongs to the one of the most interesting biogeographical regions in the world due to the high level of the endemism in fauna (Reyjol et al., 2007). Understanding of the processes behind such high number of endemic taxa is the ultimate challenge for the modern biologists. It's barely impossible to have a clear vision on the modern biodiversity patterns without knowledge of the geological processes which are behind it. Study of sediments and fossils together with investigation of modern population genetic structure allows us to estimate the approximate time of new taxa origin/speciation/age. Overlaying that data on geological story of Earth allows to provide different evolutionary hypothesis and scenarios.

Areas with high number of taxa are often called "hotspots" (Myers, 1988). Such regions rich in biological diversity (e.g. Balkan, Carpathians, Wallacea) have always been in focus of the biogeographers. One of the most recent reconsideration of biogeographical regions in the world placed Caucasus into the Palearctic region (Cox, 2001). That placement is historically stable and comes without any discussions. Great Russian zoogeographer and ichthyologists Berg L.S., in numerous works (1932, 1934, 1949) provided a scheme in biogeographical regionalization of Palearctic (Holarctic) based on the freshwater fish distribution. According to that scheme, as it was reviewed by Naseka (2010), the Caucasus region is situated in the Mediterranean Subregion of Holarctic Region and includes parts of the Ponto-Caspian-Aral Province in the Caspian and Black Sea District, neighboring Fore-Asian, Mesopotamian and Iranian Provinces (Naseka, 2010).

According to Berg, inland freshwaters of Caucasus and neighbor regions have a following structure:



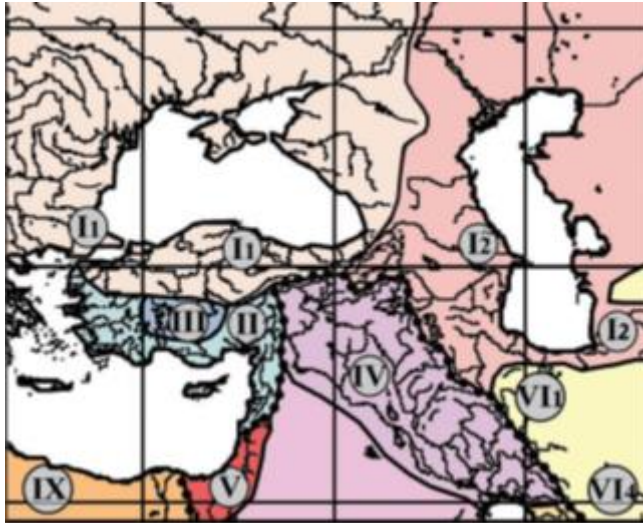


Figure 4. Scheme of zoogeographic delineation of southern-western Holarctic Region and neighbouring regions from Berg (1940; fig 20); I – Ponto-Caspian-Aral Province; I<sub>1</sub> – Black Sea District, I<sub>2</sub> – Caspian District. II – Mediterranean Province; III – Central Anatolian Province; IV – Mesopotamian Province; V – Syrian Province; VI – Iranian Province; VI<sub>1</sub> – Teheran District; VI<sub>4</sub> – Fars District; IX – African Region. (adapted from Naseka, A., 2010).

Abell and colleagues (2008) provided a map of freshwater ecoregions of the world based on distribution and composition of freshwater species. They defined ecoregion as “one or more freshwater systems with a distinct assemblage of natural freshwater communities and species” (Abell et al., 2008).

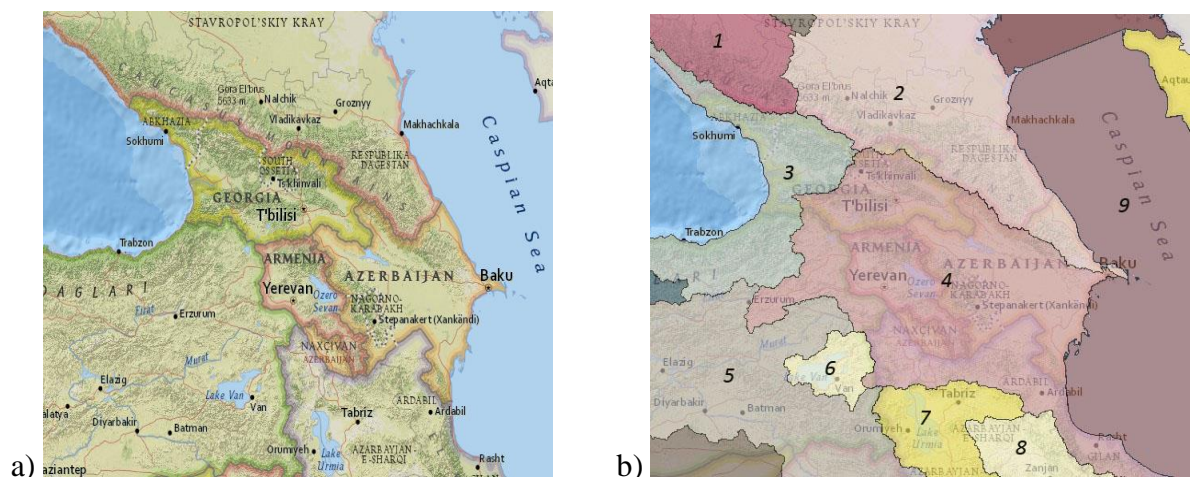


Figure 5. a) Physical map of research region with countries' borders; b) The same with ecoregions overlaid. 1) Kuban; 2) Western Caspian drainages; 3) Western Transcaucasia; 4) Kura South Caspian drainages; 5) Upper Tigris and Euphrates; 6) Lake Van; 7) Orumiye; 8) Caspian highlands; 9) Caspian Marine. Adapted from <http://www.feow.org/> (Abell et al., 2008).



According to that regionalization Azerbaijan's inland freshwaters belong to two ecoregions: almost entire country is covered by Kura South Caspian ecoregion and only small part in the north belongs to the Western Caspian drainages.

Naseka (2010) reconsidered both schemes, and in his paper he proposed following regionalization, using beta diversity index and cluster analysis of lists of freshwater fish taxa from the Caucasus region:

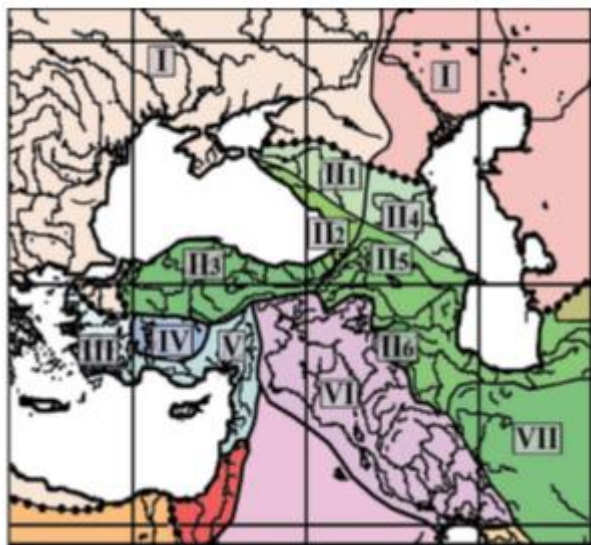


Figure 6. Scheme of zoogeographic delineation of southern-western Holarctic Region and neighbouring regions by Naseka. Holarctic region: I – Ponto-Caspian province; West Asian Region: II – Caucasian Province; II<sub>1</sub> – West Ciscaucasian District; II<sub>2</sub> – West Transcaucasian District; II<sub>3</sub> – North Anatolian District; II<sub>4</sub> – East Ciscaucasian District; II<sub>5</sub> – East Transcaucasian district; II<sub>6</sub> – Urmia District; III – West Anatolian Province; IV – Central Anatolian Province; V – South Anatolian Province; VI – Mesopotamian Province; VII – Iranian Endorheic Province. (adapted from Naseka, 2010).

Based on the following delineation Caucasus region appears much more heterogeneous as compared to the regionalization presented in the mentioned studies. Probably, drainages of Caucasus form a unique biogeographical unit. The main difference between schemes presented by Abell and Naseka relays in the separation of Caucasus to the western and eastern parts of rivers that flowing from the northern decline of Greater Caucasus mountain ridge (defined as Ciscaucasia) and flows of the southern decline (Transcaucasia). Azerbaijan is situated within Eastern Transcaucasian district, small part northward from Baku with drainages flowing directly to the Caspian Sea belongs to Eastern Ciscaucasian District. Naseka (2010) outlined the borders of East Transcaucasia: “the Kura-Aras drainage, rivers of the Lenkoranskaya (Talyshskaya) lowland flowing from the slopes of the Talyshskiy [Talysh] Ridge, and rivers in the Lesser Caucasus. The main rivers in the region include the Kura with

tributaries Araks [Aras, Araxes], Razdan [Zanga], Aragvi, Iori, Alazani, Chrami, Atstev [Akstafa], Arpa, then Vilyashchay, Lenkoran', and Safid Rud". Eastern Ciscaucasia: "includes rivers of the Caspian basin from Kuma in the north down to Sumqayitcay and some alpine lakes of the Great Caucasus. In the north, the Chornyye Zemly Desert, isolated lakes and marshes lie adjacent to the Kuma River drainage. This area also includes the Vostochnyy Manych River, which is now a partly dried and isolated drainage connected in its upper section with the Kalas River (part of the Don River drainage). The main rivers are Kuma, Terek with tributaries Gizeldon, Ardon, Uruk, Malka, Argun' and Sunzha, Sulak formed from the confluence of the rivers Avarskoye Koisu and Andiyskoye Koisu, and Samur." (Naseka, 2010).

The biogeographical patterns of the research region appear very complex and closely connected to the geological history. Durand et al., (2002) mentioned that Southwestern and Middle Asia during Late Miocene was an important center for evolution of euryhaline fauna. That was connected to the gradual closing of the Tethys Sea. In the terms of geological development of the Earth surface, we cannot separate Middle Asia from the Caucasus. Furthermore, we cannot separate the evolutionary development of any kind animal population from the processes which lead to reestablishment of land surface. During the rearrangement of landscape, animals with low vagility are captured within different de-novo formed physical barriers. Spirins, like other freshwater fishes confined to the water reservoirs inside their ecological equilibrium. Appearance of geographical barriers may have lead to separation of the original population and as a result a limitation of the gene flow and consequent allopatric speciation.

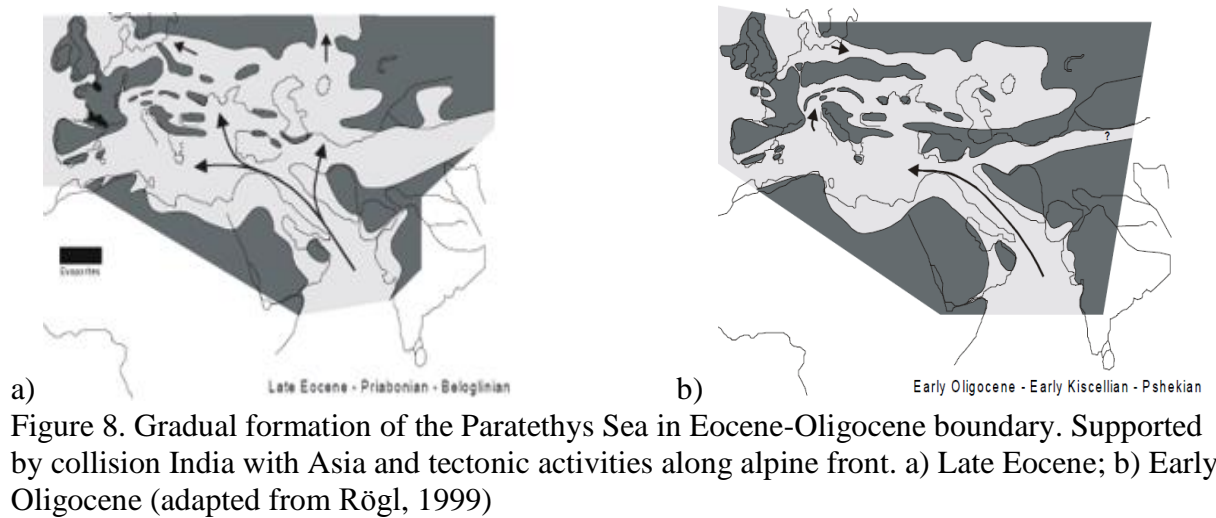
## 2.5 Geological history of Caucasus and Caspian Sea

PERIOD	EPOCH SERIES	AGE			AGES MA SIC	
		MEDITER- RANIEN	BLACK SEA	CASPIAN SEA		
QUATERNARY	HOLOCENE Q <sub>2</sub>					
	PLEISTOCENE  Q <sub>1</sub>		NEOEUXINIAN KARANGATIAN UZUNLARIAN	KVALINIAN GIRKAN KHAZARIAN	0.01 0.08 0.41	
			CHAUDIAN GURIAN	BAKUNIAN APSHERONIAN	0.70 1.0 1.80	
		PLIOCENE  N <sub>2</sub>	GELASIAN	EGRISSIAN	AKCHAGILIAN	
			PIACHENZIAN	KUYALNIKIAN		2.58
ZANCEAN	KIMMERIAN		3.0 3.60			
NEOGENE	MIOCENE  N <sub>1</sub>	MESSINIAN	PONTIAN		5.33	
		TORTONIAN	MEOTIAN		7.11	
		SERRAVALIAN	SARMATIAN		11.0	
		LANCHIAN	KONKIAN KARAGANIAN CHOKRAKIAN TARKHANIAN		13.6	
		BURDIGALIAN	KOZAKHURIAN SAKARAULIAN		16.4	
		AQUITANIAN	UPHLISTSIKHEAN		19.1	
					23.8	
	PALEOGENE	OLIGOCENE	CHATTIAN			28.0
E <sub>3</sub>		RUPELIAN  KHADUMIAN			33.7	

Figure 7. Correlation of the Upper Miocene-Pleistocene stratigraphic schemes of the Mediterranean and Black Sea-Caspian Sea regions (adapted from Adamia et al., 2011).

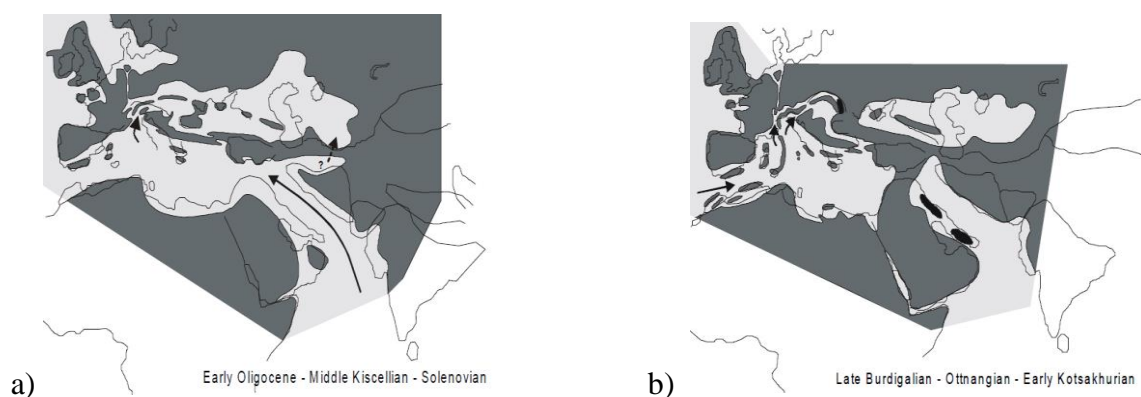
The Tethyan Realm had been reorganized during Late Eocene – Early Oligocene (Eocene-Oligocene boundary) and Caucasus had formed in a result of African/ Apulian / Arabian-Eurasian continent-continent collision. Tethys ocean was completely vanished by collision of Indian subcontinent with Eurasia. The Indian ocean and intercontinental Paratethys Sea were born. The tectonic activity along alpine front led to orogenesis processes from the Pyrenees eastwardly to Lesser Caucasus. The Paratethys Sea was separated from the

Mediterranean, creating respectively Paratethyan (northern) and Mediterranean (southern) domains as a relicts of Tethyan Realm. (Rögl, 1999, Naseka, 2010).



Term Paratethys was proposed by Laskarev (1924) outlining region from Alps to Aral Sea, separated from other remnants of Tethyan Sea by an Alpine-Caucasian uplift. In the Early Oligocene Paratethys Sea had isolated for the first time. It produced anoxic bottom conditions and high rate of endemism in that area (Rögl, 1999). At the same time and during the rest of the Oligocene, cyprinid fishes colonized Europe with the westward dispersal movement from Asia in the north of Turanian sea (modern Turanian depression) (Naseka, 2010).

Connection of Paratethys with Indian Ocean appeared repeatedly several times. (Rögl, 1999). In the terms of Caucasus formation and hydrological network around it, significant events started from the Late Burdigalian and lasted till the end of Miocene.



The Greater Caucasus, at present time is situated between Black and Caspian Sea. It's marginal region between Europe and Asia. During Early Miocene it was an island inside Paratethys Sea, referred by numerous authors as Greater Caucasus Island (Popov, 2006, Amadi, 2008). In the Early Middle Miocene (mid-Chokranian stage: 16.4 -13.6 mya (Adamia et al., 2011)) a land bridge connecting Greater Caucasus Island, modern Asia Minor (Anatolia) and Africa, appeared. Since late Miocene (mid -Sarmatian time approximately 11 mya) the Iranian and Anatolian land had been connected. (Naseka, 2010). The Caucasus river network began its formation in the Miocene. This theory is supported by that fact that first freshwater animal fossils from Caucasus are dated to the Miocene (Naseka, 2010).

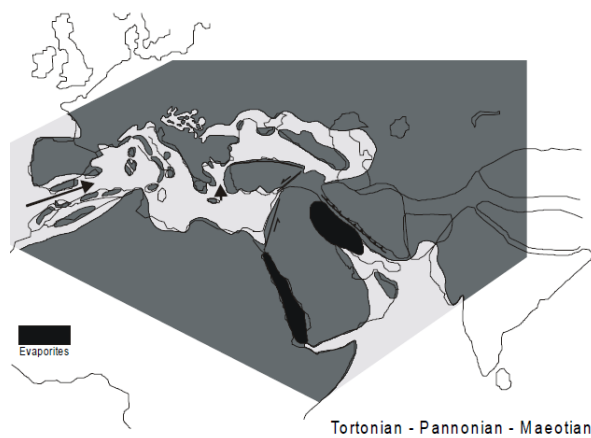


Figure 10. Final enclosure of Paratethys and formation of the Pannonian Lake. The Great Caucasus Island is still separated from the mainland. Tortonian (11 – 7 mya). (Adapted from Rögl, 1999).

During the Late Miocene the outstanding event known as Late Messenian Salinity Crisis had taken place. (Popov et al., 2006) Late Messenian Salinity Crisis at current stage of knowledge can be described by double-phase model. At first phase sea level drop was less than 100 meters and affected only Mediterranean basin margins. Second phase can be characterized by a drastic drop of sea level, up to 1500 meters and affected all Paratethyan Realm (Clauzon, 2001). Bianco (1990) pointed that in result of Messenian Salinity Crisis allowed dispersion of freshwater fish through the oligohaline Paratethys Sea formatting present endemic ichthyofauna.

Before that, at Late Tortonian - Early Messenian time Black Sea depression had already outlined modern shape, but on the north out of Greater Caucasus mountain range and Stavropol High it maintained contact with the Eastern Part of Paratethyan Realm represented by South Caspian Depression which at present time forms southern part of Caspian Sea. Kura gulf separated Greater and Lesser Caucasus.



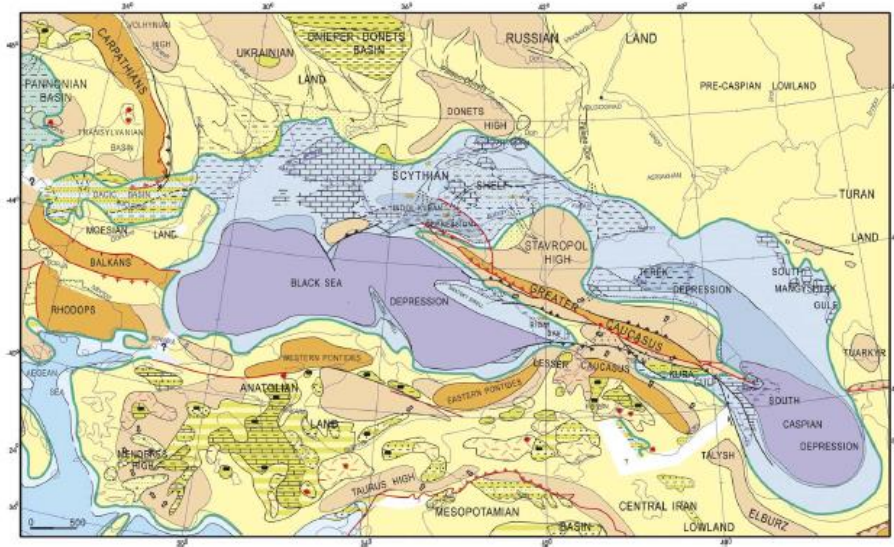


Figure 11. Research region (Eastern Paratethys Realm) in the Late Miocene (Tortonian and Messenian) on the recent geographical base. Compiled by L.B. Ilyina, I.G. Shcherba and S.O. Khondkarian, (adapted from Popov, 2004).

After the enclosure of Gibraltar straight, which caused salinity crisis and drop of the sea level, the whole Paratethyan Realm had been reorganized. The region of our interest is located in the Eastern part of realm which refers to Euxinian-Caspian basin. At Pontian time, this basin was enlarged by a transgression along its eastern and northern margins. Salinity of the basin was low, but it never dropped under 5-8 ppt (Popov et al., 2006).

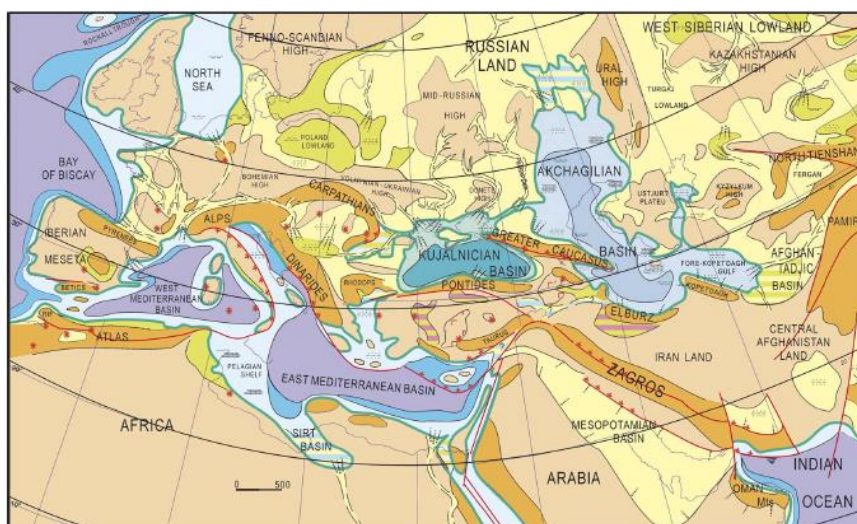


Figure 12. Palinspastic reconstruction for the middle-late Pliocene (Popov et al., 2004).

In the Pliocene (5,33 – 1,8 mya) the Ciscaucasian straight was closed separating eastern and western part of the Paratethys realm. Western part with the remains of the Panonian lake and the modern Black and Azov Seas which contributes to the Kujalnician

basin. The Caspian and Aral Seas basins with the major drainages of that time e.g. paleo-Kura, paleo-Volga, paleo-Terek, paleo-Amudarja formed the Akchagilian basin. The major regression supported by the reduction of salinity had taken place during Early Pliocene within Akchagilian basin (Popov, 2006).

Main orogenesis events in the Greater Caucasus took place in the Early Pliocene (Late Kimmerian – cca 3.6mya, see figure 6). As a result of that orogenesis Black and Caspian Sea completely outlined modern shapes (Popov, 2006).

In the Plesitocene Earth went through episodes of global climatic changes followed by rapid cooling. Those glacial events had a cyclic nature and were related to variations of the Earth orbit around the sun (the Croll-Milankovich theory;(Williams et al. 1998)). Caspian Sea in the Pleitocene had a several episodes of transgression caused by ice sheets melting in the interglacials. One of the largest transgressions had taken place in Kvalinian (cca 0.005 mya (Mamedov, 1997)).



Figure 13. Caspian Sea in the Early Kvalinian transgression. (adapted from Mamedov, 1997).

Pleistocene glaciations followed by marine transgressions and regressions probably had a great impact on the modern compound of the freshwater fishes in Azerbaijan. It can be determined by the connection of lower reaches of the rivers as a result of the strong and numerous marine transgressions. (Perea et al., 2010).

Hypothetically, the major part of Cyprinidae and Percidae have foraged in the sea and spawned in the estuaries and deltas of the Ponto-Caspian basin since Late Sarmatian time 10 – 8.5mya (Naseka, 2010). It seems likely that ancestors of spirilins at the early evolutionary stage had a connection to the brackish water, for example very close related *Alburnus* genus can also persist into lower part of drainages (Berg, 1949). In the further development of the clade and environment spirilins become restricted to the fast flowing oxygen rich water reservoirs. *Alburnoides bipunctatus* shares distribution range with some further representatives of the Leuciscinae lineage e.g. *Alburnus alburnus*, *Rutilus rutilus*, *Scardinius erythrophthalmicus* and *Squalius cephalus* and it can be explained by similar conditions that had occurs across the glacial refugees at Pleistocene glaciation. (Perea et al., 2010).



### **3 Main goals of the present study:**

- To estimate level of genetic variability inside genus *Alburnoides* in Azerbaijan using nuclear and mitochondrial markers.
- To provide barcoding data for spirilins from Azerbaijan
- To compare sequences from specimens representing geographically isolated populations.
- To provide phylogenetic analysis of genus *Alburnoides* using sequences produced during this study and to compare them to already published sequences.
- To estimate the time of divergence between lineages inside genus.
- To reconstruct biogeographical scenario.

## 4 Materials and methods

### 4.1 Samples collection

Samples were collected by electrofishing, cast net and dragnet. 268 specimens of *Alburnoides* were collected during four field trips (2012-2015). Fish were fixed and stored in 96% ethanol. The individuals included in this study originate from 29 localities in Azerbaijan and two localities in Slovakia. Specimens of *A. bipunctatus* collected in Slovakia were used mostly as outgroup for phylogenetic analyses. Localities in Azerbaijan were chosen to cover different biogeographical units inside the country. We can divide them into a three major groups – 1) Rivers of Greater Caucasus hydrological network in Northern Azerbaijan; 2) Rivers of Talysh Mountains hydrological network in Southern Azerbaijan, and 3) Rivers flowing directly to Caspian Sea on the north-east of Azerbaijan. With reference to Abell (2008) these localities belong to two ecoregions, the Kura South Caspian drainages and the Western Caspian drainages. In order to Naseka's delineation, selected localities are situated in East Ciscaucasian District and East Transcaucasian district respectively. See map with the collection points.

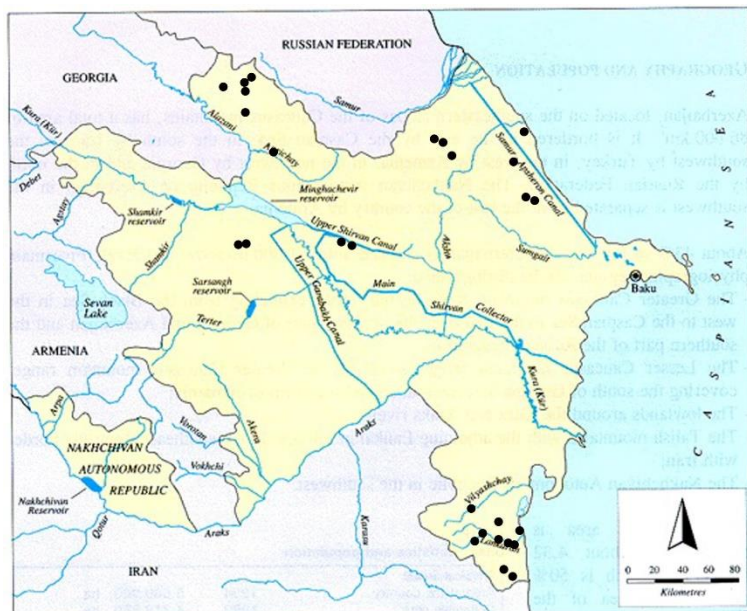


Figure 14. Map of Azerbaijan with main drainages including distribution of sampling sites.

As it was mentioned before the central region of country is not covered by sampling sites because of unsuitable ecological features for rifle minnow's persistence. Information regarding sampling sites, their geographical coordinates and name of the rivers presented in

the Table 1 in appendix.

#### **4. 2 DNA extraction and PCR**

DNA was extracted from pelvic fin clip using GenAid commercial extraction kit following manufacturer's protocol. The molecular analysis of four genes has been performed in this study (two mitochondrial – cytochrome b and cytochrome oxidase subunit I. and two nuclear – recombination activating gene I and rhodopsin gene).

I used following primers: for cytochrome b: GluF –

(5'AACCACCGTTGTATTCAACTACAA3') and ThR -

(5'ACCTCCGATCTTCGGATTACAAGACCG3') (Machordom and Doadrio, 2001) ; for

COI: Fish1F: (5'TCAACCAACCACAAAGACATTGGCAC3') and Fish1R -

(5'TAGACTTCTGGGTGGCCAAAGAATCA3'), (Ward et al., 2005); for RAG :Rag1F -

(5'AGCTGTAGTCAGTAYCACAARATG3') and Rag9R - (5'GTGTAGAGCC

AGTGRTGYTT3') (Quenouille et al., 2004); for rhodopsin: Rhod193F -

(5'CNTATGAATAYCCTCAGTACTACC3')

and Rhod1073R – (5'CCRCAGCACARCGTGGTGATCATG3') (Chen, 2003).

The polymerase chain reaction was performed on Bioer GenePro cyclers. The 25 µl of reaction mix for amplification contained: 9.7 µl of ddH<sub>2</sub>O; 12.5 µl of PPP mix; 0.65 µl of forward primer; 0.65 µl of reverse primer; 2 µl of isolated DNA.

The cycles were prepared for each gene separately. For cytochrome b: denaturation for 180 s at 94°C; 35 cycles at 92°C for 45 s; 90s at 48°C; 105 s at 72°C and final extension at 72°C for 7 min.

For cytochrome oxidase subunit I: denaturation for 10 min s at 94°C; 30 cycles at 94°C for 60 s; 90s at 58,5°C; 60 s at 72°C and final extension at 72°C for 7 min.

For recombination activating gene: denaturation for 5 min at 95°C; 31 cycles at 94°C for 40 s; 60s at 60°C; 120 s at 72°C and final extension at 95°C for 1 min.

For rhodopsin: denaturation for 5 min at 94°C; 30 cycles at 94°C for 60 s; 60s at 59°C; 90 s at 72°C and final extension at 72°C for 7 min.

Obtained PCR products were controlled by horizontal electrophoresis. Reaction mix (2µl) was applied on agarose gel (1,05g agarose in 100ml of TAE with 2 µl of EtBr) and with the reference ladder (GeneRuler™ 100bp DNA Ladder Plus for 30 minutes with voltage 110V).

PCR products were purified with ethanol precipitation following this protocol:

1) Increase the volume of reaction mixture to 100 µl by adding 75µl of ultrapure water.

- 2) Transport mix into sterile labelled microtubes and add the 10µl of 10% sodium acetate.
- 3) Add 200µl of frozen absolute ethanol and shake tubes.
- 4) Centrifuge for 10 min at 14000 rpm.
- 5) Carefully remove the supernatant, precipitated DNA forms a pellet on the bottom of microtube.
- 6) Add 100 µl of frozen 70% ethanol and centrifuge for 10 min at 14000 rpm.
- 7) Remove the supernatant and dry microtubes at 40°C for 30 min in the vacuum centrifuge.
- 8) Add 25µl of ultrapure water and store samples in freezer.

The quality and relative quantity of DNA after purification were controlled on spectrophotometer MaestroGen MaestroNano.

Sequencing was performed in Macrogen Service Centre (Amsterdam, Netherlands).

Sequences were aligned using ClustalW algorithm (Thompson, 1994) in Geneious software (Geneious version 9.1.2 (<http://www.geneious.com>, Kearse et al., 2012) and compared to published sequences using BLAST tool. Published sequences were downloaded from <http://www.ncbi.nlm.nih.gov> to provide more objective and full phylogenetic analysis (41 sequences of cytochrome b; 50 sequences of COI and 31 of RAGI).

### 4.3 Dataset preparation

#### 4.3.1 Datasets for MP, ML and Bayesian methods

Different number of sequences was obtained for each gene. The cytochrome b has the highest number of successfully isolated, amplified, purified and sequenced samples – 80 sequences. There are 60 sequences of *Alburnoides* genus representatives between them. 20 sequences that remain are from different close related taxa (e.g. genera *Alburnus*, *Rutilus*). They were included into dataset because sampled individuals were hardly distinguishable from spirilins especially on the juvenile stage. I further aimed to obtain barcodes of fish from the region, which is genetically still underexplored. Barcodes should serve for the potential future research. The samples of other genera were also used as outgroups in the phylogenetic analyses. The dataset was further enriched by adding 42 downloaded sequences from the GenBank database. The total alignment of 122 sequences of cytochrome b has been normalized to 1140bp length and used in the analyses. I produced 32 sequences of COI and added 50 downloaded sequences for dataset (Geiger et al., 2014, Roudbar et al., 2016), the length of alignment – 701bp. The 23 sequences of RAGI were produced and 31 were added to

dataset with total number of replicates – 54 and length 1438bp. For rhodopsin gene there are only few sequences in public access, so we used only our data for analysis. The alignment was cut to the 888bp length. Datasets were converted to NEXUS, PHYLIP and MEG format.

#### 4.3.2 Haplotype network dataset

Smaller special dataset was created for haplotype network. It includes only *Alburnoides* genus representatives. It contains 58 sequences of cytochrome b normalized to 895bp length. Specimens represented localities from all republic.

#### 4.3.3 Concatenation of genes

Finally, some sequences were concatenated if two or more genes were available from one individual. Concatenation of genes was provided in Geneious software ver. 9.1.2. (Kearse et al., 2012). Three datasets were created. One of them included 34 sequences with total length 4037 bp. The sequences of same genes from some other cyprinid fishes which are closest relatives (e.g. *Tinca*, *Phoxinus*, *Leuciscus*) were included to this dataset. It was used for MP and Bayesian analyses to provide comparative phylogeny of *Alburnoides* genus and their closest relatives. The next two datasets were prepared for Molecular Clock estimations and were smaller due to the high sensitivity of the BEAST analysis. First one included 12 sequences with total length 3968bp, it included downloaded sequences for different cyprinid fishes (e.g. *Cyprinus carpio*, *Scardinius erythrophthalmus*, *Chondrostoma nasus*, *Tinca tinca*, *Phoxinus phoxinus*, *Leuciscus leuciscus*). This dataset was made to estimate the time of divergence for *Alburnoides* lineage. Second one included sequences strictly from Azerbaijan to estimate time of divergence of separate clades inside genus. It included 24 sequences of mitochondrial genes (cyt b and COI) with total length 1765bp.

#### 4.4 Model testing

For the statistical selection of the best-fit model of nucleotide substitution for the Bayesian and ML analyses jModelTest2 was used (Guindon & Gascuel, 2003; Darriba et al., 2012) which using parallel computing. After computing the likelihood scores the best-fit model was tested under AIC (Akaike information criterion) (Akaike, 1976) for each gene.

For cyt b the suggested model was:

Lset base = (0.2864 0.3079 0.1360 ) nst=6 rmat=(0.5700 16.8935 0.5124 1.8061 7.1076)  
rates=gamma shape=1.7090 ncat=4 pinvar=0.5580;

For COI:

Lset base=(0.2600 0.2829 0.1745 ) nst=6 rmat=(1.0000 22.1683 1.0000 1.0000 11.8406)  
rates=gamma shape=0.8140 ncat=4 pinvar=0.5230;

For RAGI:

Lset base=equal nst=6 rmat=(1.0000 2.7619 1.7060 1.7060 4.1189) rates=gamma  
shape=0.7310 ncat=4 pinvar=0.7210;

For Rhod:

Lset base=(0.1768 0.3329 0.2462 ) nst=6 rmat=(1.0000 5.6299 1.0000 1.0000 3.3352)  
rates=equal pinvar=0.7970

#### **4.5 Maximum likelihood method**

The PhyML version 3.0 (Guindon et al., 2010) was used for maximum likelihood estimation. This package provides a reduced computing time comparing to other maximum-likelihood packages with simple hill climbing algorithm that able to modify tree topology and branch length at one time (Guindon et al., 2010). Input data were converted to PHYLIP format using Geneious ver. 9.1.2. (Kearse et al., 2012). Maximum likelihood estimation was done for each four genes separately.

#### **4.6 Maximum parsimony**

Maximum parsimony method relies in constructing phylogenetic trees under maximum parsimony criterion (Nei & Kumar, S. 2000). The MEGA software (MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets (Kumar, Stecher & Tamura 2015) was used to establish MP analysis in this study. It appeared user friendly, enough accurate and less time consuming comparing to other software for MP tree construction (e.g. PAUP).

#### **4.7 Bayesian analysis**

Bayesian Markov chain Monte Carlo (MCMC) methods become a common tool to provide phylogenetic and evolutionary models after their implementation in the statistical phylogeny (Mau & Newton, 1997; Yang & Rannala, 1997) and invention of easy-to-use software, such as MrBayes (Huelsenbeck & Ronquist, 2001). MrBayes version 3.2 (Huelsenbeck & Ronquist, 2003) was used for estimation posterior probabilities of phylogenetic trees. The NEXUS files were created in order to results of model testing. The parameters of model were set (nst=6 rates=invgamma) for cytochrome b, RAG and COI what

corresponds to GTR model with gamma distribution. The HKY model with equal distribution (nst=6 rates=equal) was set for rhodopsin gene. The analysis was done with number of generation 5,000,000 with the sample frequency every 10,000 trees.

#### **4.8 Haplotype network**

Haplotype networks were produced in PopART software (Population Analysis with Reticulate Trees) <http://popart.otago.ac.nz> using median joining method that combines Kruskal's algorithm and Farris's maximum-parsimony heuristic algorithm (Bandelt et al., 1999). Two haplotypes networks were created including all cytochrome b haplotypes from Azerbaijan and samples only from northern part of republic (see Dataset for more details). File in NEXUS format was used as an input file. The obtained haplotype network was manually redrawn in Adobe Photoshop Cs3.

#### **4.9 Molecular clock method**

The BEAST (Bayesian Evolutionary Analysis by Sampling Trees) and BEAUti (Bayesian Evolutionary Analysis Utility) software version 1.7 (Drummond et al., 2012) were used for molecular clock method that implements MCMC algorithms for Bayesian phylogenetic inference, coalescent analysis and divergence time dating (Drummond et al., 2012). XML file was generated in BEAUti from NEXUS file including concatenated sequences. The modeltest results for interspecific concatenate were:

For cytochrome b:

Lset base=(0.3019 0.3109 0.1249 ) nst=6 rmat=(1.0000 15.0477 1.0000 1.0000 9.9860)  
rates=gamma shape=1.9540 ncat=4 pinvar=0.5590; what corresponds to GTR+I substitution model;

For COI:

Lset base=(0.2761 0.2682 0.1483 ) nst=6 rmat=(1.0000 13.1634 1.0000 1.0000 8.0252)  
rates=gamma shape=1.1430 ncat=4 pinvar=0.5880; what corresponds to GTR+I substitution model;

For RAG1:

Lset base=equal nst=6 rmat=(1.6619 4.4633 1.6619 1.0000 5.9798) rates=gamma  
shape=0.0970 ncat=4 pinvar=0; what corresponds to GTR substitution model;

For Rhod:

Lset base=(0.1795 0.3302 0.2487 ) nst=2 tratio=2.1914 rates=equal pinvar=0.8060; what corresponds to HKY substitution model.

For interspecific concatenate uncorrelated relaxed clock (Drummond & Rambaut, 2006) was chosen as most suitable for study objectives. The tree-prior was set to Speciation: Birth-Death Process (Gernhard, 2008). The time of hypothetical cladogenesis or most recent common ancestor was taken from published data (Wang et al., 2012) – 12.02mya with standard deviation – 0.63 with normal prior distribution. The analysis was performed number of generation 50,000,000.

For intraspecific concatenate which included sequences of mitochondrial genes another run of model test was performed. For both genes HKY model was suggested. The tree prior was set to Coalescent: Constant size (Kingman, 1982).



## 5 Results

The main goal of this study is to describe genetic variability within the genus *Alburnoides* from Azerbaijan using mitochondrial and nuclear markers. Gene sequences were obtained from specimens from geographically and ecologically different regions in order to perform phylogenetic analyses with subsequent biogeographical interpretations. My goal is to illustrate phylogenetic relationships within and among populations and bring an insight into evolutionary and biogeographical history of the genus.

My hypothesis is that population from one biogeographical region would have more similar haplotypes than populations from different biogeographical regions. For example, individuals from the Greater Caucasus hydrological network and from the Talysh Mountains river basin(s) would be genetically distant, i.e. would form separate clades in the phylogenetic tree, in case my hypothesis was corroborated. Phylogenetic relationships within the genus *Alburnoides* are not in accord with the proposed hypothesis as the observed pattern cannot be explained only by simple geographical isolation.

There are four main distinguishable clades in the obtained phylogenetic trees within the genus *Alburnoides*. These clades represent logical units and serve for better interpretation and easier understanding of the results:

1) EichN+S – large clade formed by samples of *Alburnoides eichwaldii* from both northern (Greater Caucasus) and southern (Talysh Mountains) populations. Individuals from the southern populations occur sympatrically (at the same localities) with members of the EichS clade, originally undistinguishable in the field based on morphology and overall appearance.

2) EichS – clade formed by samples of *Alburnoides eichwaldii* strictly from the southern population, i.e. rivers of Talysh Mountains hydrological network (e.g. Lenkaran river).

3) Cis – clade formed by samples from East Ciscaucasian region (Naseka, 2010) (see fig 5). Members of this clade represent putatively undescribed species named as *Alburnoides* sp 5. This clade also includes samples from Russia, namely *Alburnoides kubanicus* and *A. fasciatus* from Kuban region.

4) Eur – European clade formed by individuals of *Alburnoides bipunctatus* from the Danube drainage (Slovakia), together with sequences of *A. bipunctatus* from different parts of Europe (France and Germany) available in public access in GenBank databases. Further species, subspecies or undescribed lineages belonging to this clade: *Alburnoides bipunctatus strymonicus*, *A. thessalicus*, *A. cf. prespensis*, *A. ohridanus* from Balkan region including

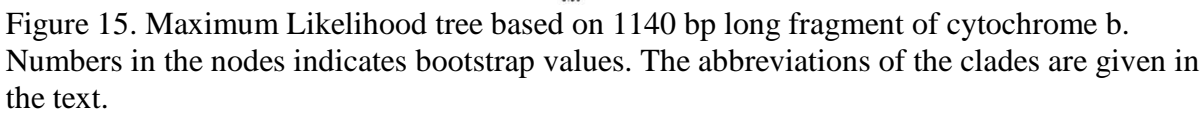
Albania and Greece and *A. sp 2*, *A. sp 3*, *A. tzanevi* from Bulgaria and Romania, *A. bipunctatus rossicus* from Russia and Ukraine. See Appendix 3.

Four datasets were created for different genes (cytb, COI, RAGI, rhod) in this study. In addition to that, three datasets of concatenated genes were created and 1 dataset was created for haplotype network. Sixteen analyses were performed and 16 phylogenetic trees are presented.

### **5.1 Genetic variability of the genus *Alburnoides* inferred from mitochondrial cytochrome b analysis**

Dataset for analysis using cytochrome b included 122 sequences (downloaded and generated within this study) normalized to 1140 bp length. Average pairwise identity of the sequences was 92,1% and the alignment included 62,1% of identical sites. The GTR+I+G model was suggested for Maximum Likelihood and Bayesian analysis.

Generally, there is a large congruence among the results from the different analyses. Below, I summarize the support for the particular clades based on the cytochrome b data set: Eur – was relatively well supported in ML and Bayesian analyses (96 likelihood ratio, 0.968 posterior probability, in MP analyses Eur clade has an uncertain topology). Cis– wasn't well supported in ML, Bayesian and MP analyses (88 likelihood ratio, 0.774 posterior probability, less than 50% bootstrap respectively). EichS – was well supported in ML, Bayesian and MP analyses (100 likelihood ratio, 1.00 posterior probability, 99% bootstrap respectively) includes sequences from Iran (Seifali, M. et al., 2012) from the rivers very close to Azerbaijan borders. EichN+S – was well supported in ML, Bayesian and MP analyses (100 likelihood ratio, 1.00 posterior probability, 89% bootstrap respectively). The problematics of Eich N+S clade is in persistence of two inner clades that are supported by high ratios in ML, MP and Bayesian analyses. Two inner clades are formed by samples both from North and South and possible explanation for the following distribution will be discussed.



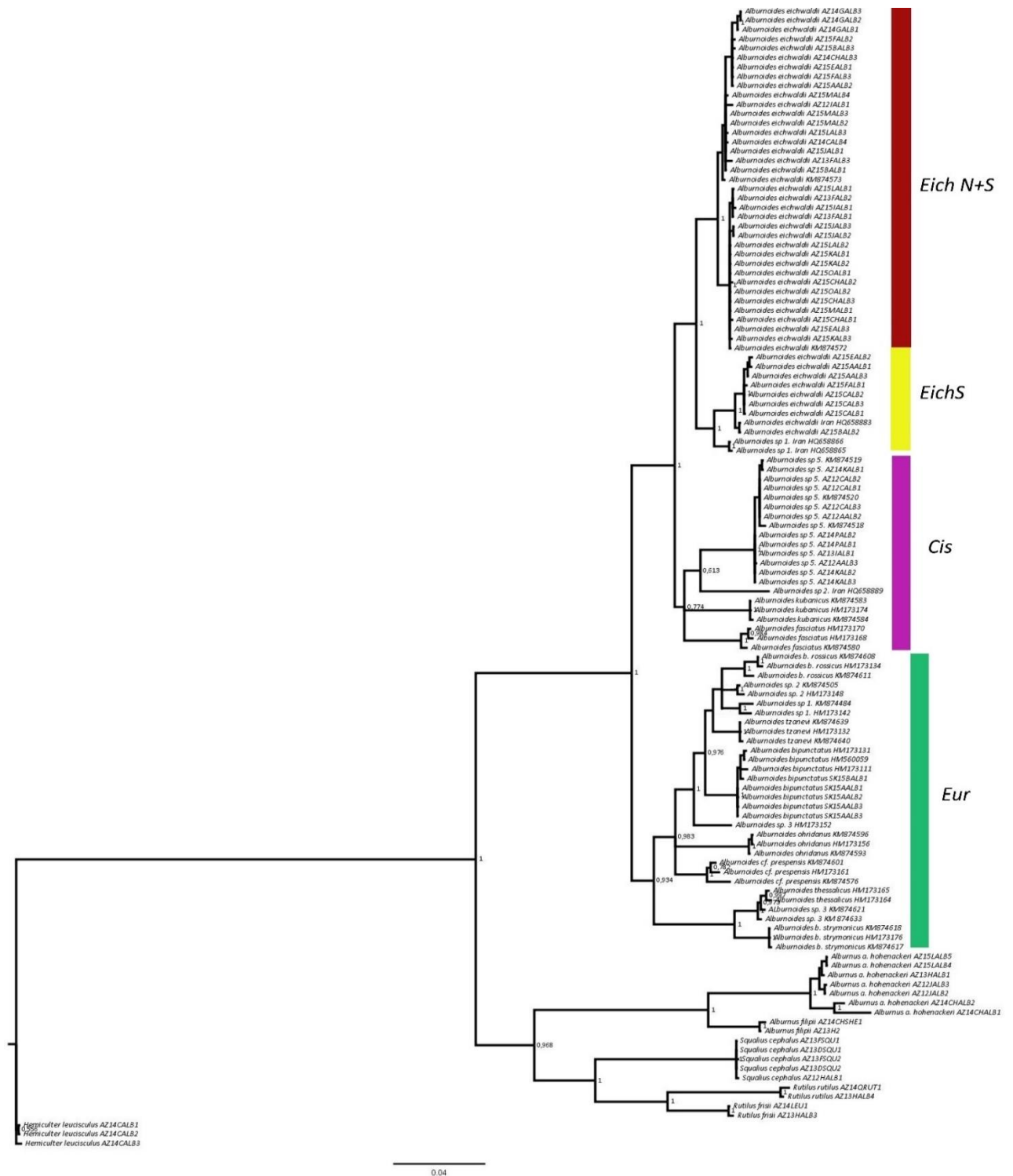


Figure 16. Bayesian tree based on the 1140bp long fragment of cytochrome b. The tree was performed for 5 million generations; the last 2500 trees were used. The numbers in the nodes indicates Bayesian posterior probabilities. The abbreviations of the clades are given in the text. *Hemiculter leucisculus* was used as an outgroup.

There are several minor differences in the MP tree topology comparing to ML and Bayesian tree. The Eur clade has another topology, *A. cf. prespensis* has another position and placed more closely to Cis clade. Like in ML and Bayesian trees nodes between *A.*

bipunctatus and *A. thessalicus*, *A. b. strymonicus*, *A. ohridanus* have very weak support. Combining these species in the one logical clade supported by low genetic variability inside it. Remaining clades have a comparable topology with other trees.

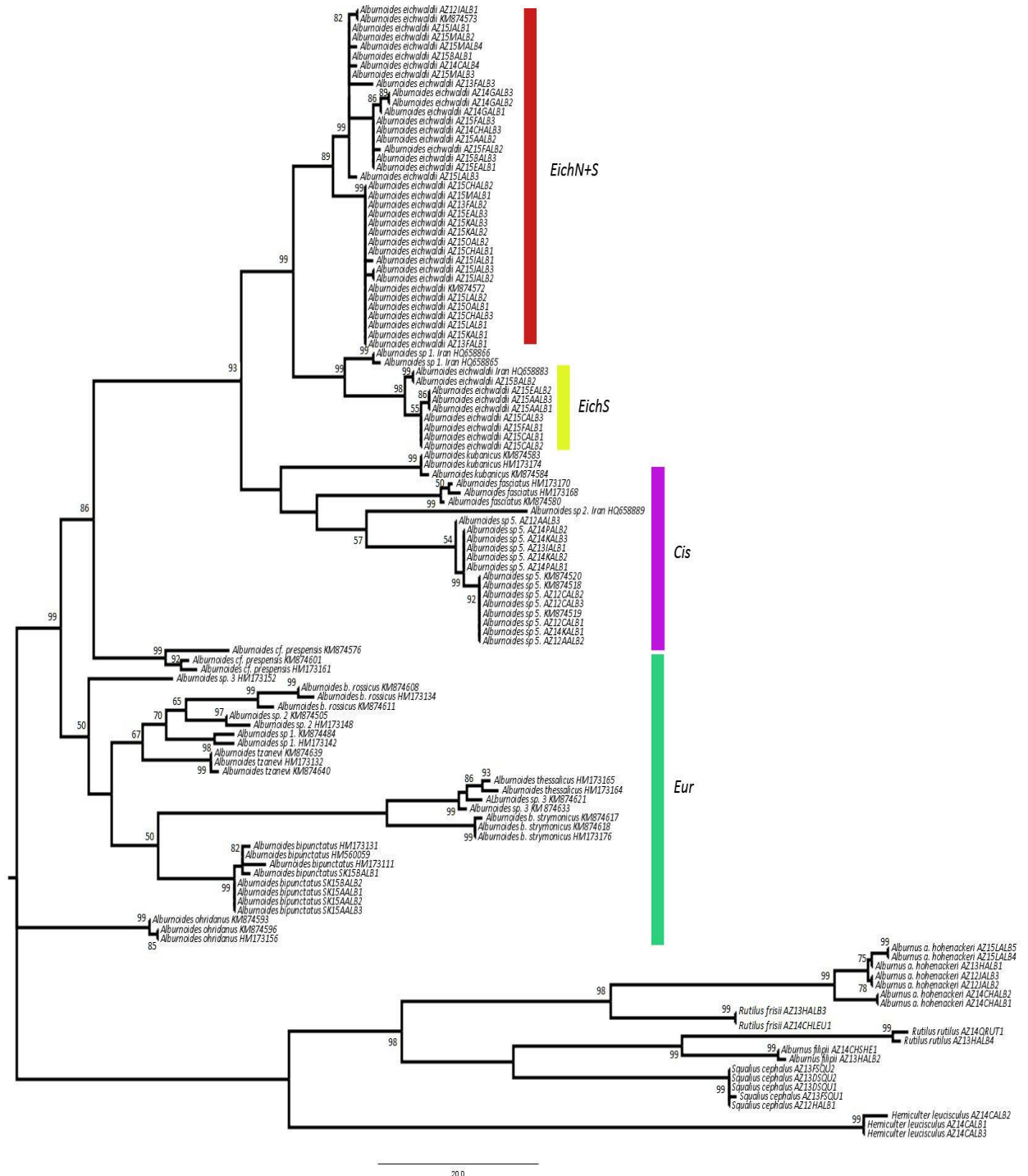


Figure 17. Maximum Parsimony analysis result based on the cytochrome b data set. The most parsimonious tree with length = 674 is shown. The consistency index is (0.453172), the retention index is (0.910947), and the composite index is 0.421685 (0.412816) for all sites and parsimony-informative sites (in parentheses). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to

the branches (Felsenstein, 1985). The values under 50 are not shown. The MP tree was obtained using the Subtree-Pruning-Regrafting (SPR) algorithm (Nei & Kumar, 2000) with search level 1 in which the initial trees were obtained by the random addition of sequences (10 replicates). The analysis involved 122 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 648 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Kumar et al., 2016). (Description of the figure was adapted from the MEGA7 software).

## **5.2 Genetic variability of genus *Alburnoides* inferred from the analysis of mitochondrial cytochrome oxidase subunit I gene.**

The analyses were performed on COI gene alignment of 82 sequences normalized to 701 bp.. For more slight visualization I left them separate. I have added sequences of COI gene from several species from Iran, which were available on the public database GenBank. These sequences contribute to a several newly described species and form a complicated topology. Between those species are: *Alburnoides qanati*, *A. coadi*, *A. sp HRE2016*, *A. nicolausi*, *A. idignensis*, *A. tabarestanensis*, *A. namaki*, *A. holciki* (Roudbar, J. et al., 2016). Summary of the clades:

Eur – was poorly supported in ML, MP and Bayesian analyses (61 likelihood ratio, 0.563 posterior probability, 53% bootstrap respectively).

Cis – was well supported in ML, Bayesian and MP analyses (98 likelihood ratio, 0.996 posterior probability, 98% bootstrap respectively).

EichS (together with Iranian sequences) – was well supported in ML, MP and Bayesian analyses (99 likelihood ratio, 0.997 posterior probability, 95% bootstrap respectively).

EichN+S – was relatively well supported in ML, Bayesian and MP analyses (99 likelihood ratio, 0.939 posterior probability, 84% bootstrap respectively).

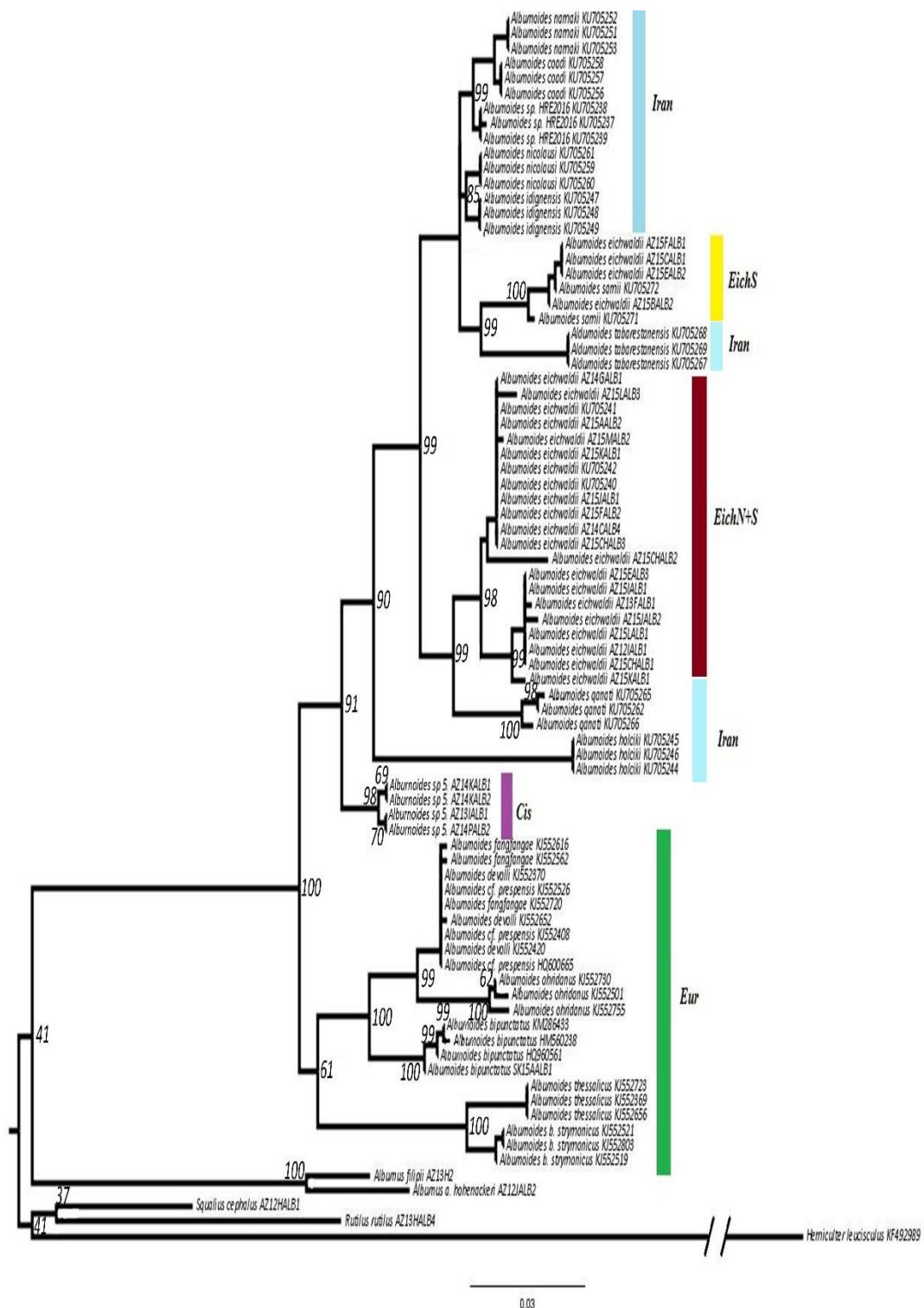


Figure 18. Maximum Likelihood tree based on 701 bp long fragment of cytochrome oxidase subunit I gene. Numbers in the nodes indicates bootstrap values. The abbreviations of the clades are given in the text. Downloaded sequences of Iranian species are highlighted with blue color.







clades are given in the text. *Hemiculter leucisculus* was used as an outgroup. Downloaded sequences of Iranian species are highlighted with blue color.

Maximum Parsimony tree has a comparable topology to Bayesian tree.

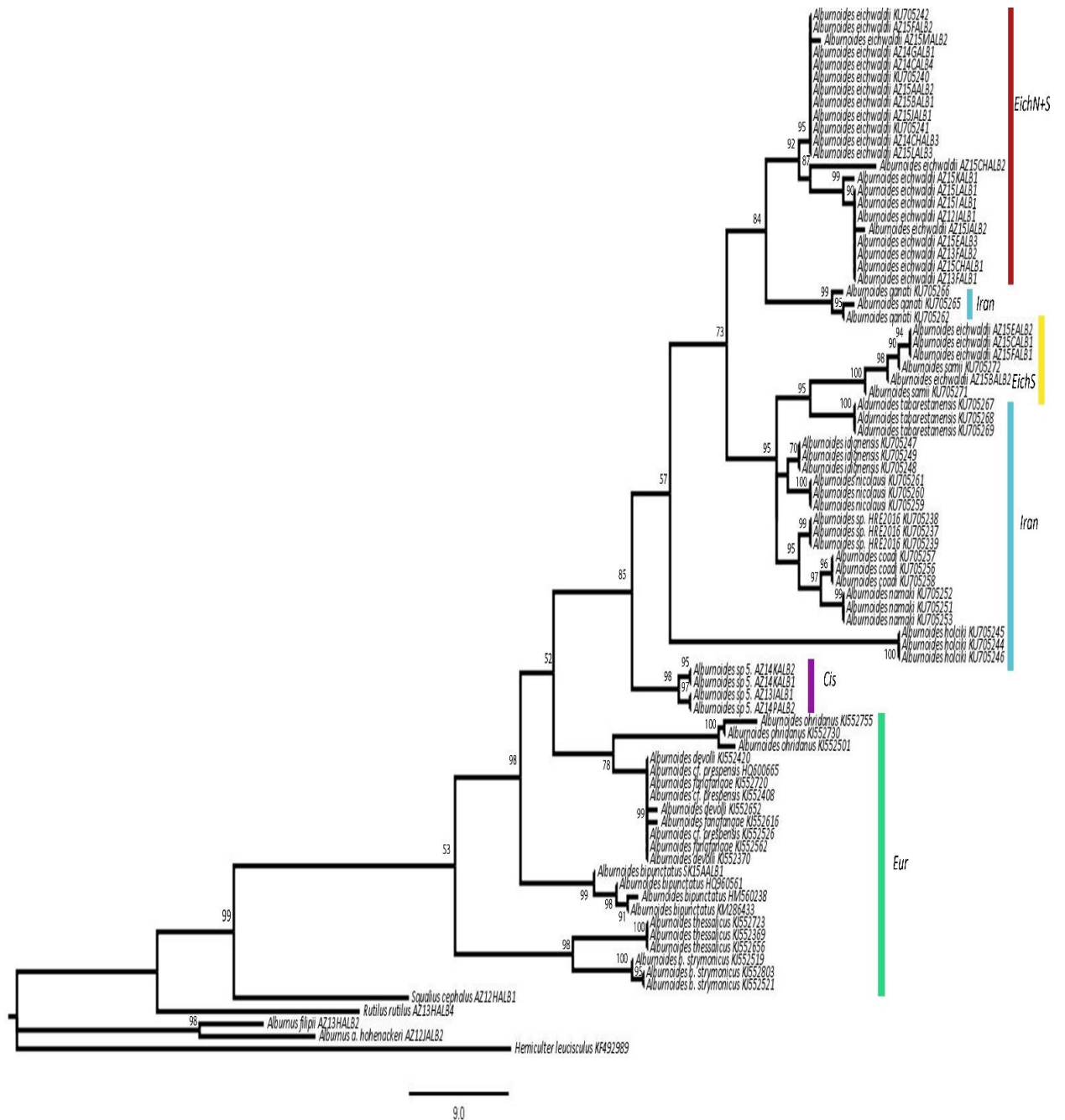


Figure 20. Maximum Parsimony analysis result based on the COI data set. The most parsimonious tree (length = 325) is shown. The consistency index is (0,498258), the retention index is (0,878481), and the composite index is 0,489246 (0,437710) for all sites and parsimony-informative sites (in parentheses). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches (Felsenstein, 1985). The MP tree was obtained using the Subtree-Pruning-Regrafting

(SPR) algorithm (Nei & Kumar, 2000) with search level 1 in which the initial trees were obtained by the random addition of sequences (10 replicates). The analysis involved 82 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 410 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Kumar, 2016). (Description of the figure was adapted from MEGA7 software). Downloaded sequences of Iranian species are highlighted with blue color.

### **5.3 Genetic variability of genus *Alburnoides* inferred from the analysis of nuclear recombination activating gene I**

The dataset included 54 sequences normalized to 1438 bp length. Obtained signals from the nuclear gene were significantly different from the mitochondrial ones. The topology is much more simple and nodes are very poorly supported. All clades that were mentioned before (Eur, Cis, EichS and EichN+S) aren't well supported; all lineages ascend to a single polytomy node, which is supported by the bootstrap support of 93 in the ML analysis and posterior probability of 0.9 in the Bayesian analysis. Topologies of three obtained trees from ML, MP and Bayesian analyses are congruent. They differ in position of *A. kubanicus*. Two internal clades of the *Alburnoides eichwaldi* are not present in the RAG1 data set, as individuals from the EichS clade (based on the mitochondrial genes) completely disperse within the EichN+S clade herein.



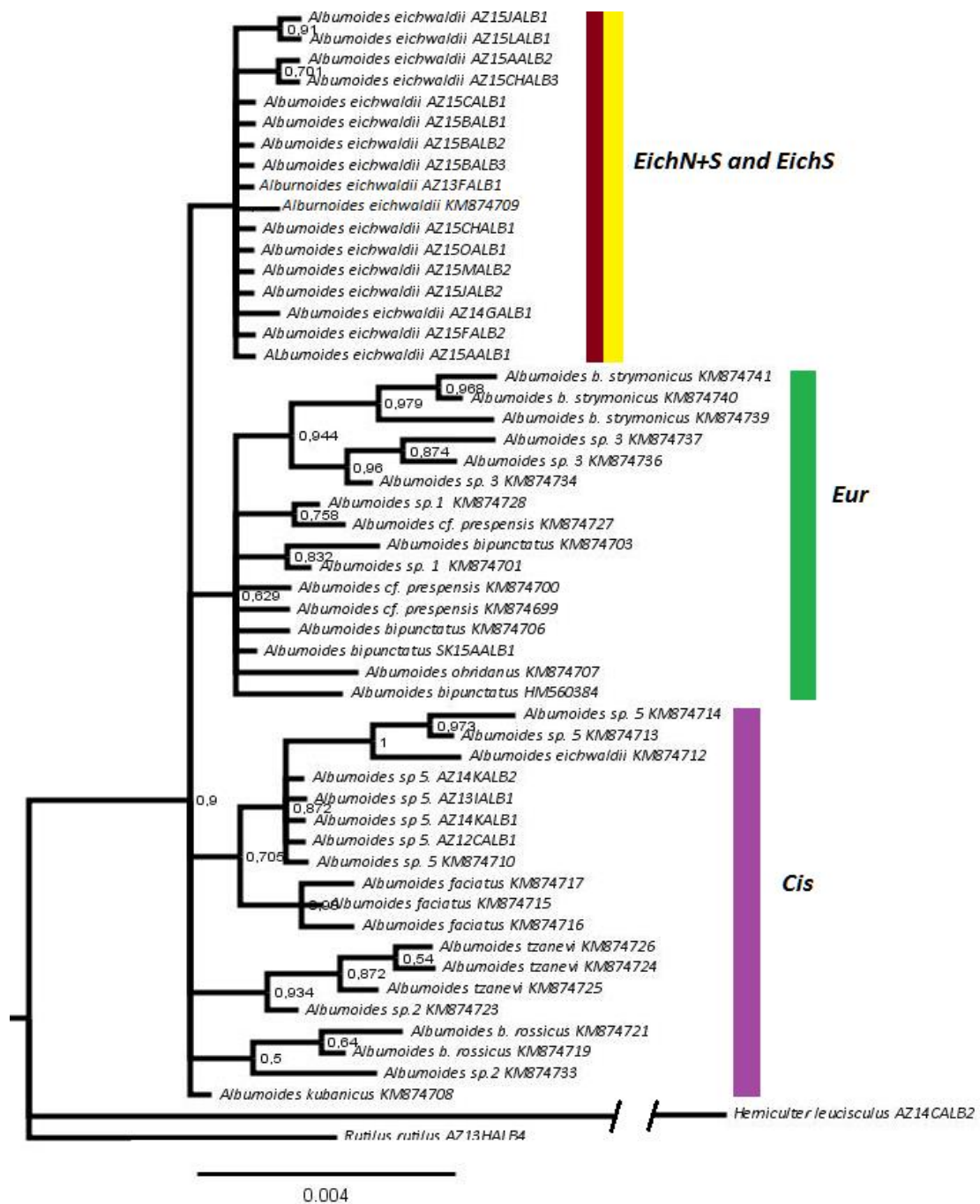


Figure 22. Bayesian tree based on 1438 bp long fragment of RAG1. The tree was performed for 5000000 generations; the last 2500 trees were used. The numbers in the nodes indicates Bayesian posterior probabilities. The abbreviations of the clades are given in the text. *Hemiculter leuciscus* was used as an outgroup.

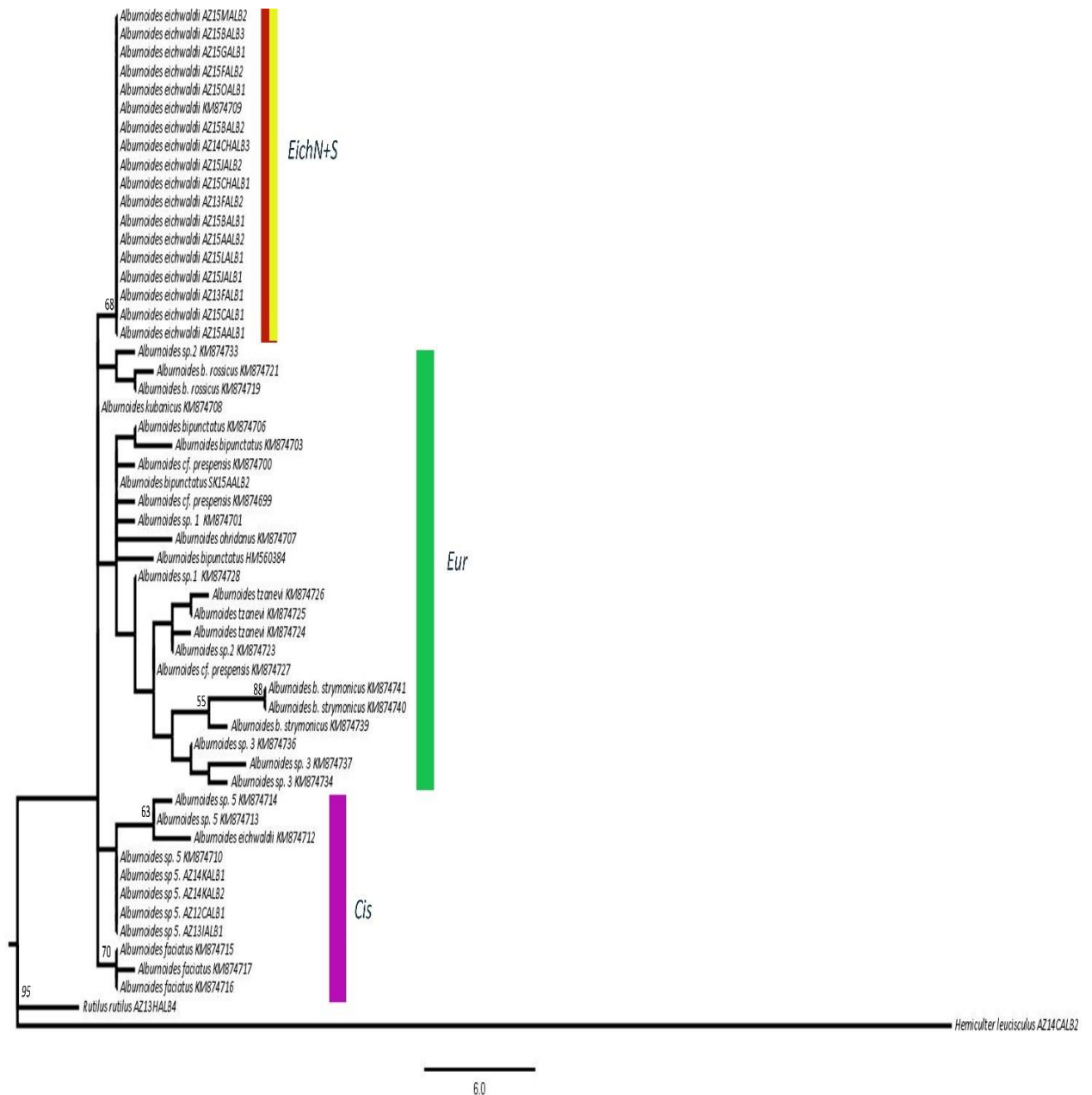


Figure 23. Maximum Parsimony analysis result based on the RAGI data set. The most parsimonious tree with length = 101 is shown. The consistency index is (0.583333), the retention index is (0.825581), and the composite index is 0.702970 (0.481589) for all sites and parsimony-informative sites (in parentheses). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches (Felsenstein, 1985). The MP tree was obtained using the Subtree-Pruning-Regrafting (SPR) algorithm (Nei & Kumar, 2000) with search level 1 in which the initial trees were obtained by the random addition of sequences (10 replicates). The analysis involved 54 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 1093 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Kumar, 2016). (Description of the figure was adapted from MEGA7 software).

#### 5.4 Genetic variability of genus *Alburnoides* inferred from the analysis of nuclear rhodopsin gene.

The smallest dataset was used for analyses inferred from the rhodopsin gene fragment. It included only 20 sequences with length of 888bp. The distribution of replicates is comparable to the results of the RAGI analyses. There are no distinguishable clades supported by significant values in the trees. Nodes inside genus are uncertain and poorly supported. *Alburnoides bipunctatus* forms a separate branch from the rest of the genus, but that node has a very low values. The *Alburnus a. hohenackeri* has a closer position to spirilins from the Azerbaijan than Eurasian spirilin (*A. bipunctatus*) sampled in Slovakia.

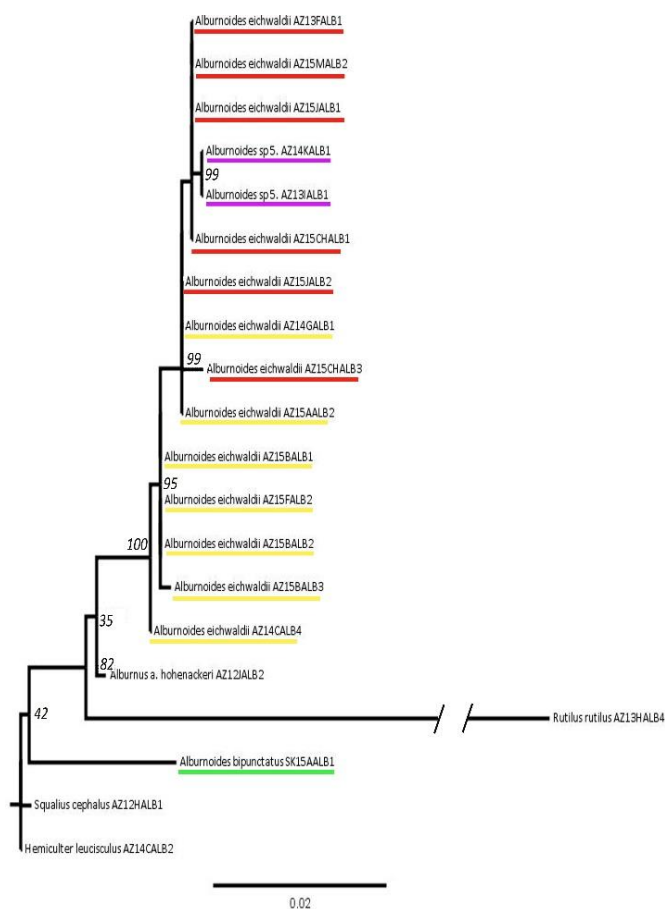


Figure 24. Maximum Likelihood tree based on 888 bp long fragment of rhodopsin gene. Numbers in the nodes indicates likelihood ratios. The abbreviations of the clades are given in the text. Colors contribute to the clades present in previous trees (green – Eur clade; violet – Cis; yellow – EichS; red – EichN+S).

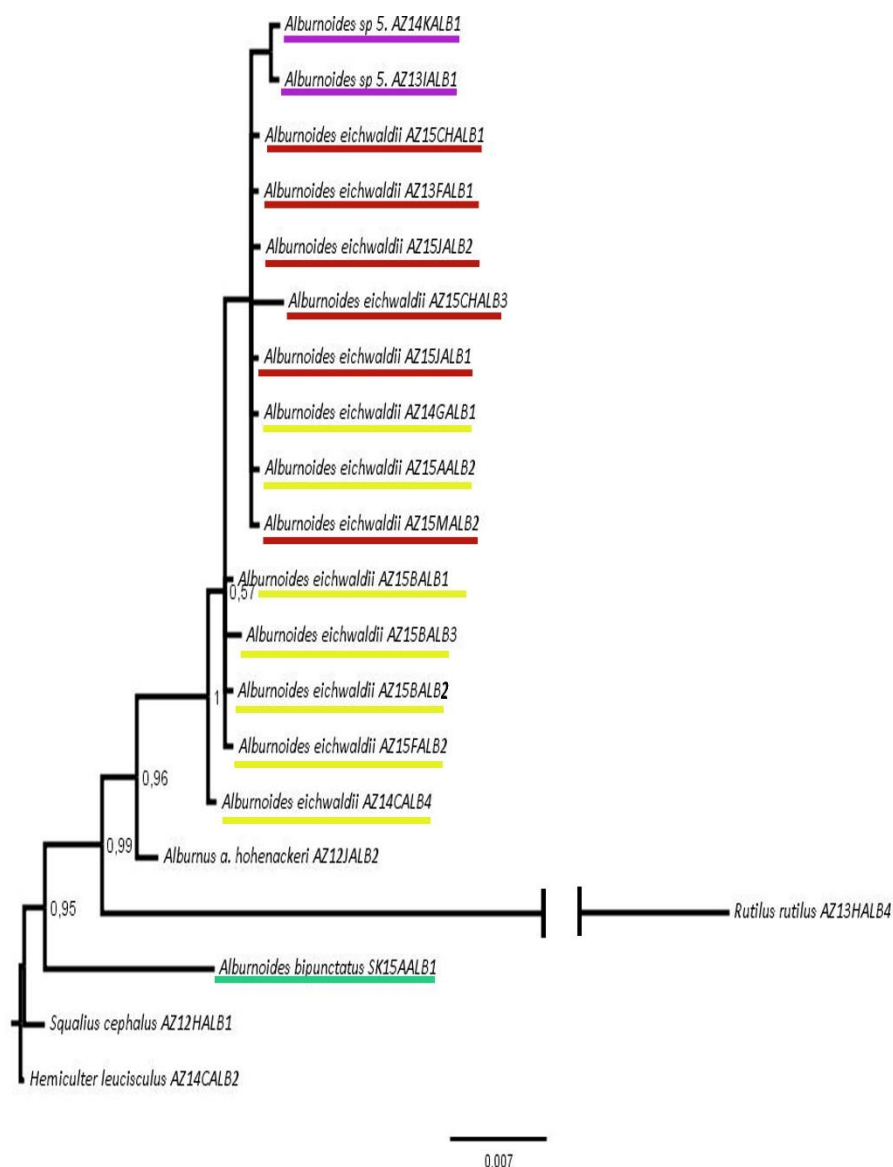


Figure 25. Bayesian tree based on 888 bp long fragment of rhodopsin gene. The tree was performed for 5000000 generations; the last 2500 trees were used. The numbers in the nodes indicates Bayesian posterior probabilities. The abbreviations of the clades are given in the text. Colors contribute to the clades present in previous trees (green – Eur clade; violet – Cis; yellow – EichS; red – EichN+S). *Hemiculter leucisculus* was used as outgroup.



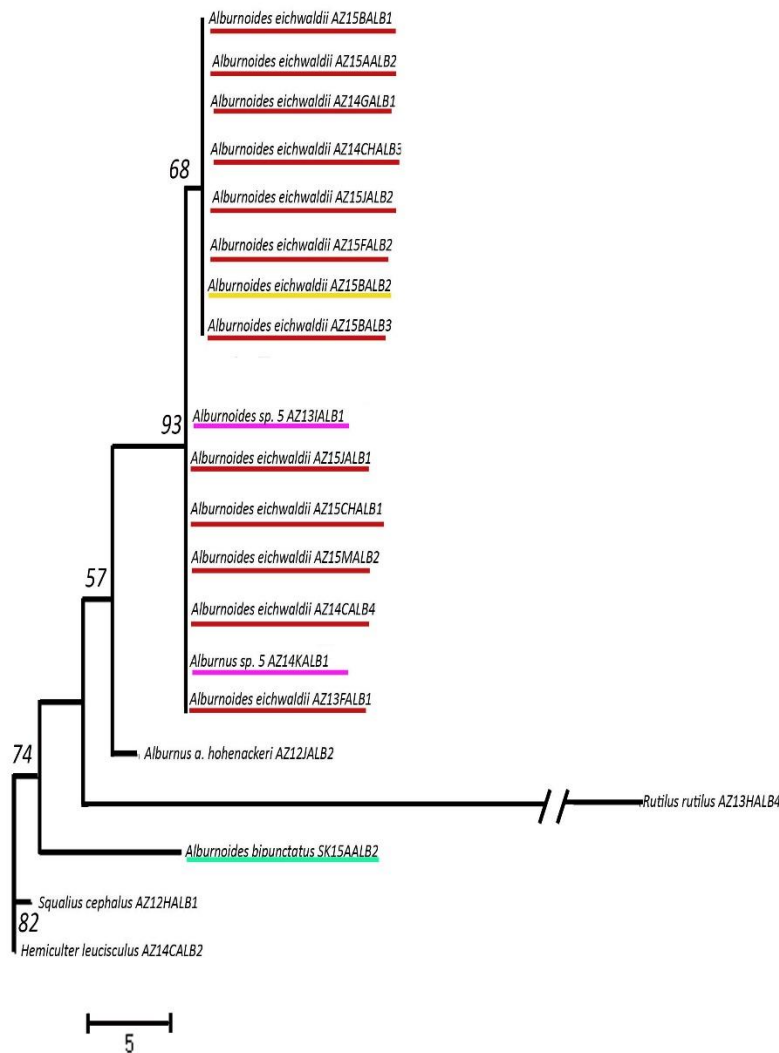


Figure 26. The most parsimonious trees based on the rhodopsin data set (length = 82) is shown. The consistency index is (0.882353), the retention index is (0.937500), and the composite index is 0.914634 (0.827206) for all sites and parsimony-informative sites (in parentheses). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches (Felsenstein, 1985). The MP tree was obtained using the Subtree-Pruning-Regrafting (SPR) algorithm (Nei & Kumar, 2000) with search level 1 in which the initial trees were obtained by the random addition of sequences (10 replicates). The tree is drawn to scale, with branch lengths calculated using the average pathway method and are in the units of the number of changes over the whole sequence. The analysis involved 20 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 763 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (Kumar, 2016). Colors contribute to the clades present in previous trees (green – Eur clade; violet – Cis; yellow – EichS; red – EichN+S). (Description of the figure was adapted from MEGA7 software).

### 5.5 The analysis of concatenated mitochondrial and nuclear genes

The results obtained from Bayesian analysis of the concatenated genes are in agreement with previously commented results and trees topologies. For concatenation I used



only self-generated sequences of the genus *Alburnoides* and several downloaded sequences of 4 genes for the taxa that were used as outgroups, namely *Leuciscus*, *Phoxinus*, *Tinca*.

The Eur clade represented only specimen *A. bipunctatus* SK15ALAB1 from Slovakia, the node is well supported by posterior probability 1.00. The Cis and EichS, EichN+S clades are also supported by same value of posterior probability. Inside the EichN+S probability the two unnamed clades formed by samples both from northern and southern localities also found great support.

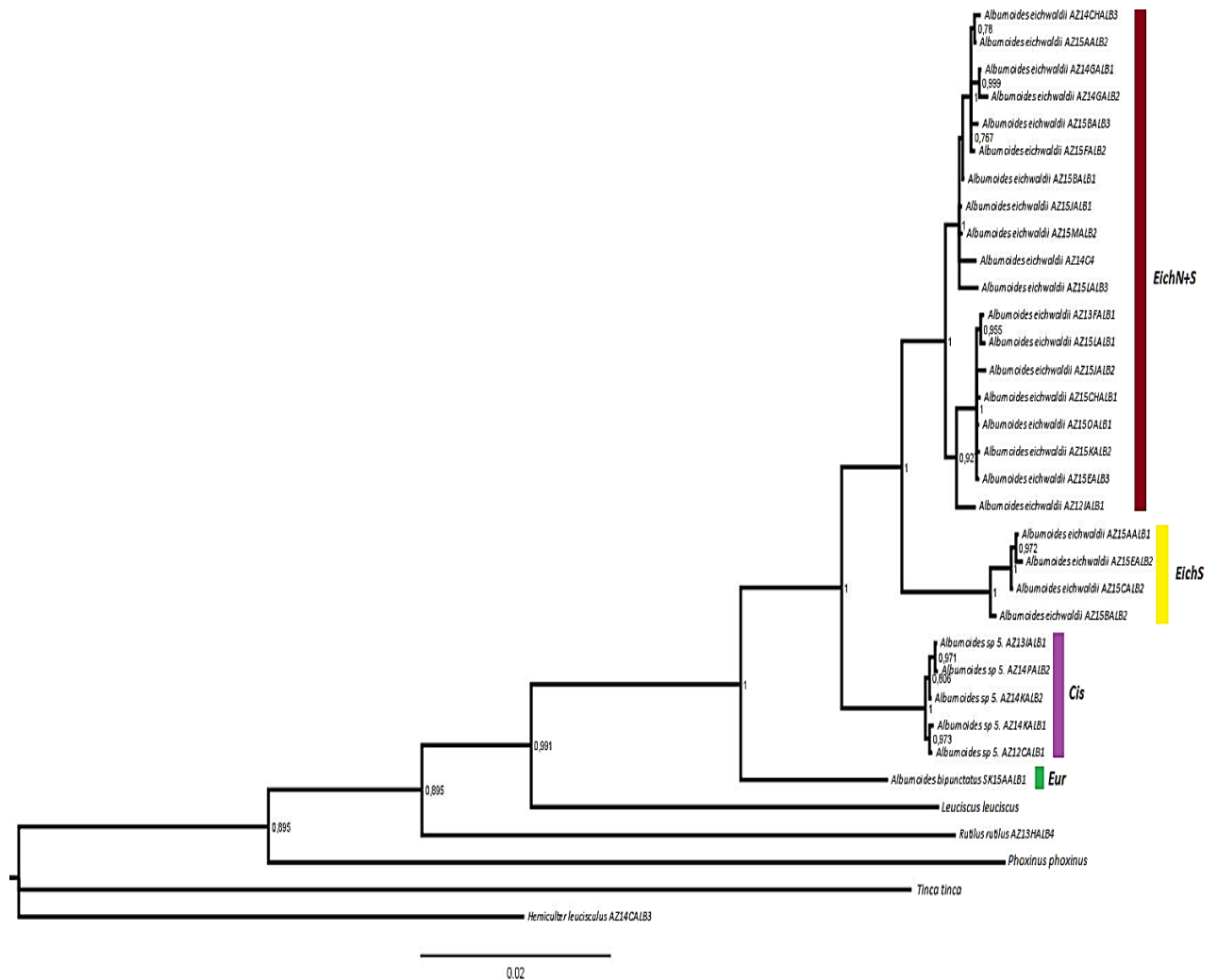


Figure 27. Bayesian tree based on 4037 bp long concatenated sequences from 4 genes (cytb, COI, RAG1, rhod). The tree was performed for 5000000 generations; the last 2500 trees were used. The numbers in the nodes indicates Bayesian posterior probabilities. The abbreviations of the clades are given in the text. *Hemiculter leucisculus* was used as outgroup.

In the both Bayesian and the MP trees upper part of EichN+S clade is formed by sequences strictly from the South (Talysh Mountains hydrological network).



Figure 28. Maximum Parsimony analysis of taxa. The most parsimonious tree (length = 492) is shown. The consistency index is (0.604534), the retention index is (0.739635), and the composite index is 0.503613 (0.447135) for all sites and parsimony-informative sites (in parentheses). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches (Felsenstein, J., 1985).

The MP tree was obtained using the Subtree-Pruning-Regrafting (SPR) algorithm (Nei, M., & Kumar, S., 2000) with search level 1 in which the initial trees were obtained by the random addition of sequences (10 replicates). The analysis involved 34 nucleotide sequences. All positions containing gaps and missing data were eliminated. (Description of the figure was adapted from MEGA7 software). Tips labels in the EichN+S clade annotated in the order to their geographical contribution (north or south). The upper part of EichN+S with samples from the south highlighted by yellow box.

## 5.6 Molecular clock method

Smaller “interspecific” dataset was created for the molecular clock calibration and estimation. It included 12 sequences 3968 bp long, two of them were from *Alburnoides* (Cis and EichS clades), the rest were from the different Cyprinidae lineages. The age estimate of the node of *Leuciscus*, *Phoxinus* and *Rutilus* genera was taken from published data (Wang et al., 2012) – and the node was dated as 12.02 mya (million years ago) with standard deviation of 0.63. For molecular clock method the Speciation: Birth-Death Process tree prior was used (Gernhard, T., 2008). The age of the node between *Alburnoides* sp 5. AZ13IALB1 and *A. eichwaldii* AZ15BALB2 was estimated as 2.35 million years old with standard deviation – 0.3.

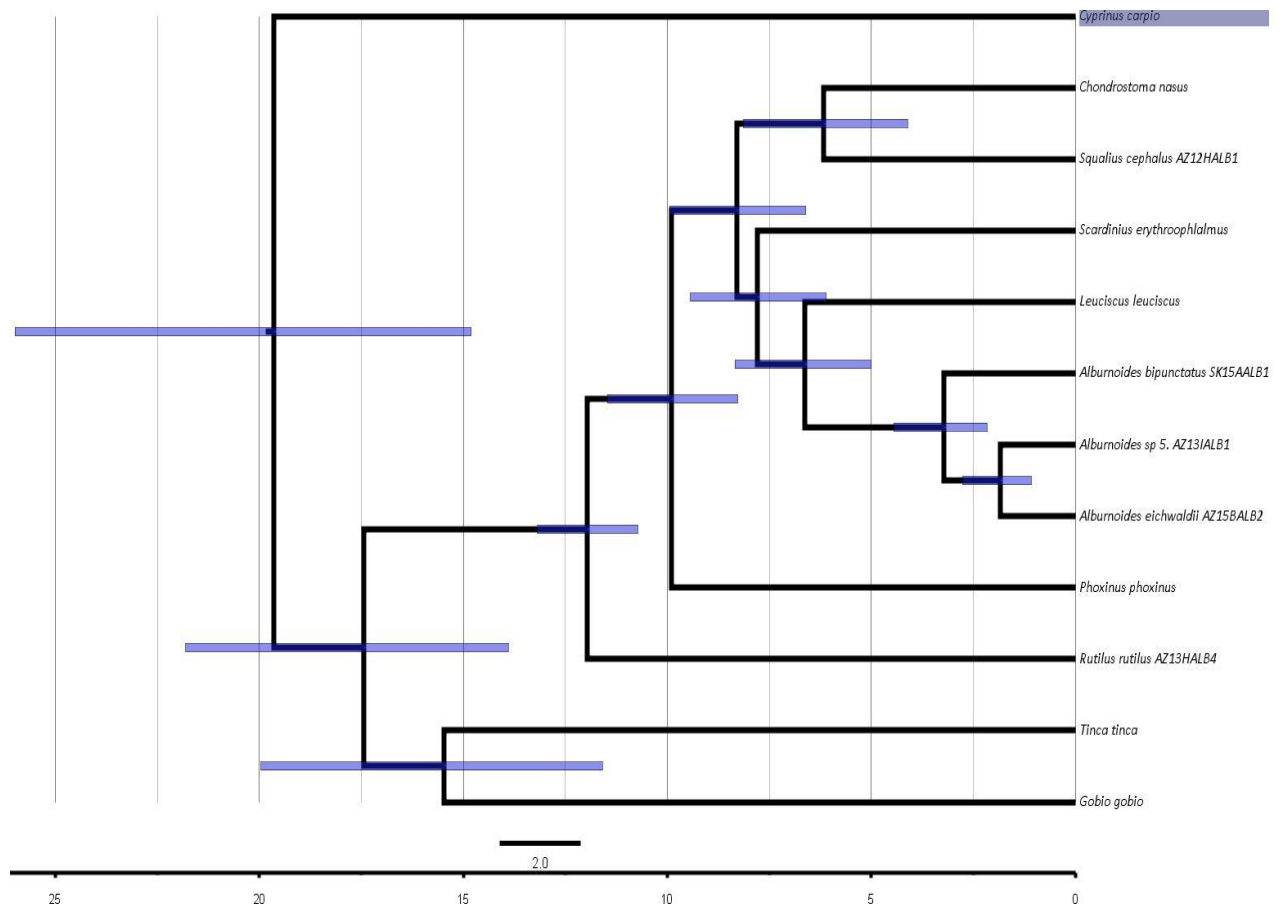


Figure 29. BEAST tree based on 3968 bp long fragment of concatenated (rhod, RAGI, cyt b, COI) genes. Two independent runs were performed and the data were combined. The tree was performed for 100'000'000 generations; the last 25000000 trees were used. The tree overlays time axis. *Cyprinus carpio* was used as an outgroup.

For the intraspecific estimation of nodes ages the analysis of 24 replicates were performed using mitochondrial cyt b and COI genes. Samples used for analysis originated strictly from Azerbaijan. The node between EichS and EichN+S clade has been estimated to split about 1.5 mya. Node inside the EichN+S clade was estimated to split about 0.56 mya.

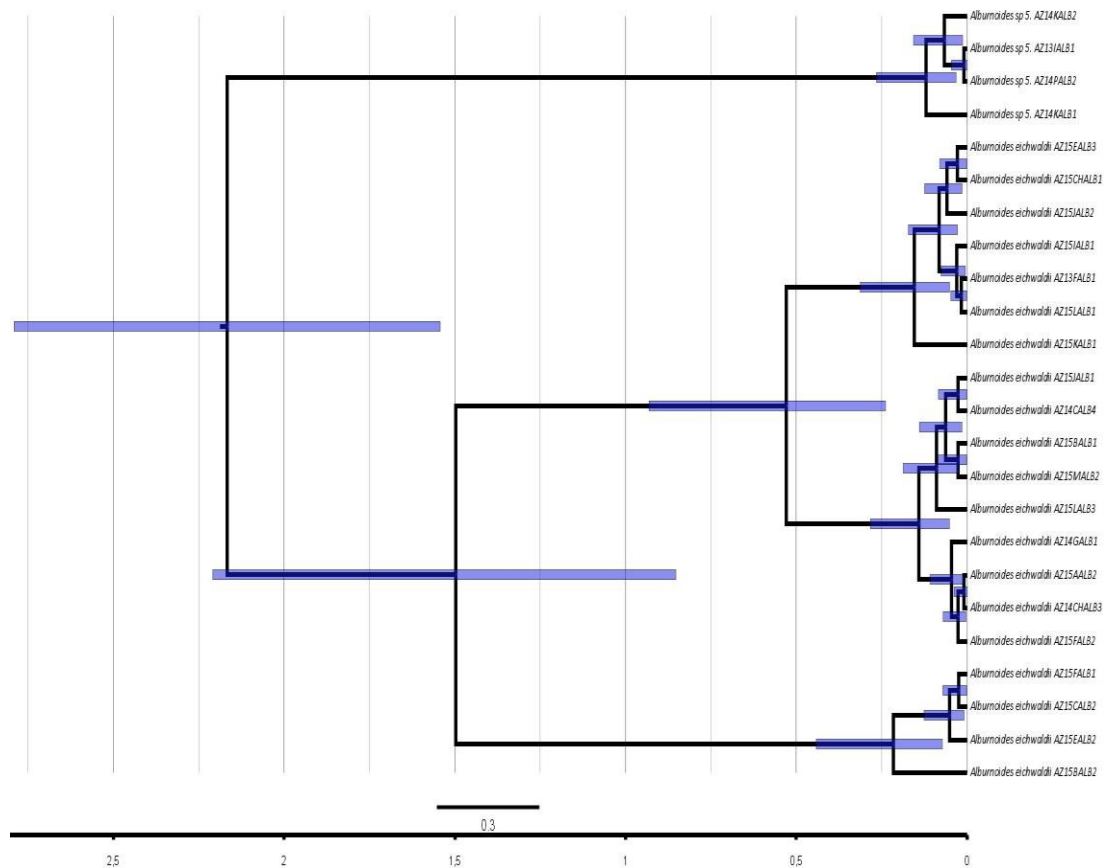


Figure 30. Figure 24. BEAST tree based on 1765 bp long fragment of concatenated COI and cyt b genes. Two runs were performed and data was combined. The tree was performed for 100000000 generations; the last 25000000 trees were used. The tree overlays time axis.

## 5.7 Haplotype network

Haplotype network shows that haplotype distribution is to some extent congruent with the geographical contribution of the samples. However, some haplotypes typical for Greater Caucasus hydrological network (North) can be found in the rivers on the South (Talysh Mountains).

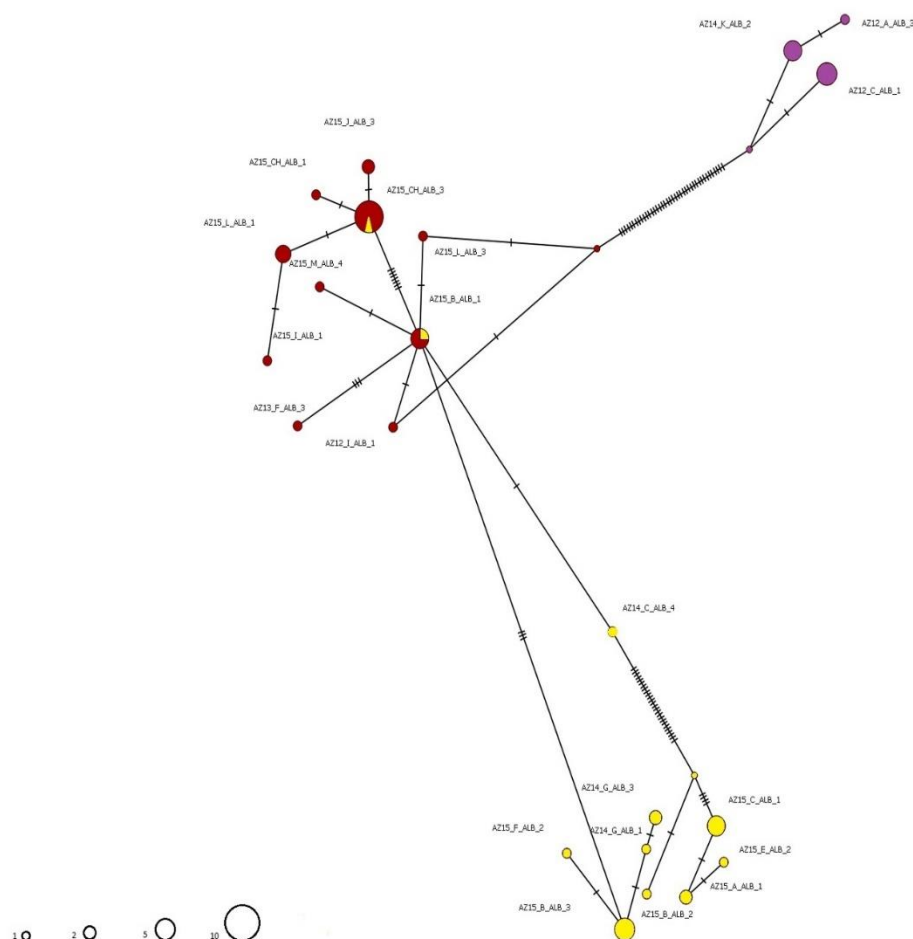


Figure 31. Median joining haplotype network based on 53 sequences of cytochrome b 895 bp long. Colors representing geographical contribution of samples. Yellow – samples collected on the South; Red – samples from Greater Caucasus hydrological network and center of republic (Kura drainage); Violet – samples from East Ciscaucasian district. The size of the circles corresponds to number of samples sharing same haplotype.

## 6 Discussion

Obtained results confirm great genetic variability inside the country. This variability was supported by almost all kinds of phylogenetic models established in this study. All clades seem to be well supported in models inferred from mitochondrial signals. The nuclear markers are more conservative and did not have enough signal to form distinguishable clades. However, concatenated dataset supports division of the four major clades. There is a lack of articles devoted to molecular phylogeny of genus *Alburnoides* particularly in Azerbaijan, which could serve for comparison with the results of this thesis. Nevertheless, there are numerous studies focused on the intrageneric relations of *Alburnoides* in Europe (Mendel et al, 2010, Stierandova, 2015) or neighboring Iran (Seifali et al., 2012, Roudbar et al., 2016). Sequences from those studies were included in the data sets of this thesis for comparison of Azerbaijan spirilins and providing jointed phylogeny. The paper by Stierandova et al., (2015) focused on the multilocus assessment of mitochondrial and nuclear sequence data and it has the highest comparative value for this thesis, mostly because the same genetic markers (cyt b and RAGI) were used and it is also one of the most recent studies. Furthermore, some specimens of *Alburnoides* sp. 5 included in the study of Steirandova et al. (2015) were collected in 2012 by Miroslav Švátora, and therefore, there is even a slight overlap in the original localities of the samples, even though their study did not focused on the Eastern Mediterranean region. Samples present both in mine and the published study correspond to the localities AZ12A and AZ12C (Tugchay river). Similarly, some specimens of *A. eichwaldii* from Kurakchay river AZ12I (see Appendix) have also been used in both studies. The small overlap in samples from the same localities allowed me to make assumptions about the accuracy of this study and to compare of results of both studies.

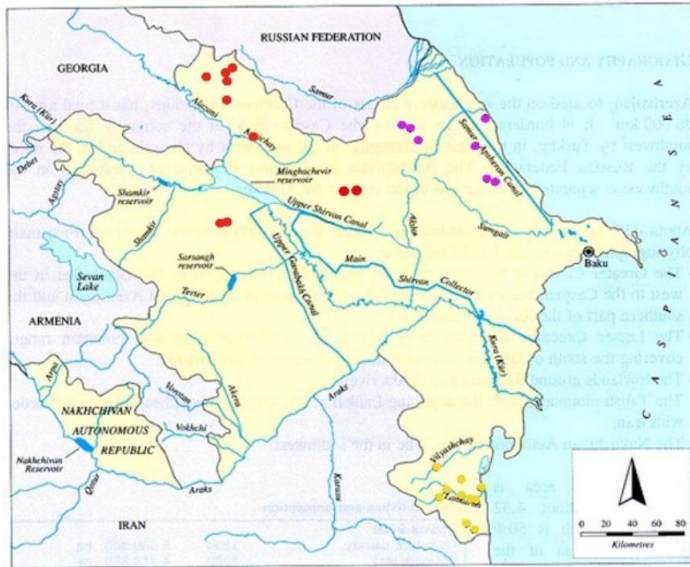


Figure 32. Map of the sampling sites. Colors correspond to the clades. (violet – Cis clade; yellow – EichS clade; red – EichN+S clade).

## 6.1 Discussion of cytochrome b signals

Species outside *Alburnoides* genus, which were included to analysis (*Squalius cephalus*, *Rutilus frisii*, *Rutilus rutilus*, *Alburnus filippii*, *Alburnus alburnus hohenackeri*) form a basal branches which are well supported (See Fig. 15, 16, 17). Those samples were included mostly for the barcoding purposes. Many nodes inside the European clade are very poorly supported and are in agreement with the published data (Stierandova, S. et al., 2015).

Likelihood ratio for the node between *Alburnoides* cf. *prespensis* and *A. ohridanus* equals 34; posterior probability lower than 0.5 and node has less than 50% bootstrap. In the published phylogeny *A. ohridanus* forms a poorly supported lineage either. Furthermore, node between *A. bipunctatus* and rest of the clade (*A. tzanevi*, *A. sp 1*, *A. sp 2*, *A. b. rossicus*) has a relative low likelihood value – 88, posterior probability – 0.976, which can be considered as low and insignificant. Similar situation occurs in discussed study where same node has a posterior probability 0.76 and all the rest values are not shown probably due to their insignificance (Stierandova et al., 2015).

The undescribed species *A. sp 1* and *A. sp 2* appear several times in mine cytochrome b trees as a part of different clades. Sequences of those species were downloaded from GenBank ([www.ncbi.nlm.nih.gov/genbank/](http://www.ncbi.nlm.nih.gov/genbank/)) and their authorship is different. Vouchers of *A. sp 1* with accession numbers HM173142 and KM874484 and *A. sp2* with accession numbers HM173148 and KM874484 were uploaded to the GenBank by Mendel and Stierandova respectively (Mendel et al., 2010; Stierandova et al., 2015). *Alburnoides sp 1*. was sampled in

the streams of Bosnia and Herzegovina and Croatia. But *A. sp. 1* with accession numbers HQ658866, HQ658865 that appears in the EichS clade (see figures 15, 16, 17) and form a very well supported independent branch (likelihood ratio = 100; posterior probability – 1.00; 99% bootstrap). Those cytochrome b sequences were submitted to GenBank by Seifali (2012) are from Rūdkhāneh-ye Kaslān stream in Iran which flows into southern part of Caspian Sea (Seifali et al. 2012).

With the *Alburnoides sp. 2* situation is similar. Two sequences from Europe (Romania and Serbia) have an authorship by Mendel and Stierandova respectively. They appear inside European clade and form a node apart from *A. bipunctatus rossicus* with a low likelihood bootstrap support – 39. In the Bayesian and Maximum Parsimony this part of European clade has another topology. *A. sp. 2* forms a node with *A. tzanevi* (posterior probability lower 0.5 and 67% bootstrap). The other sequence named *A. sp 2* (accession number HM658889) has an Seifali authorship and also has an Iranian origin namely in Rūdkhāneh-ye Āb Parrān river which flows into the eastern part of southern Caspian Sea. This specimen has an interesting position - it belongs to the Cis clade (see Fig. 16). That clade is also formed by vouchers (*A. fasciatus*; *A. tzanevi*) from Russia and our samples from northern part of Azerbaijan. Appearance of voucher from Iran in the clade formed by fish collected in the north of Greater Caucasus range is surprise and further study should be performed to confirm the results. The node between Cis clade and rest of the genus (part of Ponto-Caspian clade *sensu* Stierandova) has a similar phylogenetic signals in both studies. The rest of my dataset is not covered in Stierandova's study. The problematics of Eich N+S clade is in persistence of two inner clades that are supported by high ratios in ML, MP and Bayesian analyses. Two inner clades are formed by samples both from North and South and possible explanation for the following distribution will be discussed.



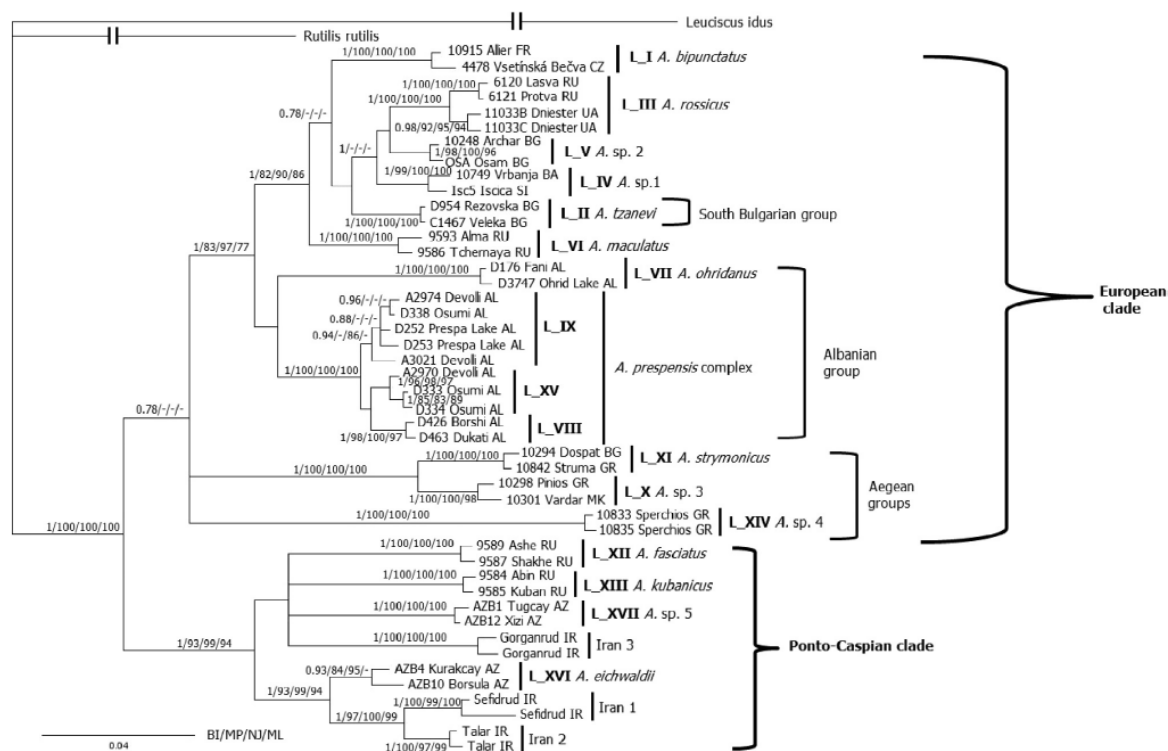


Figure 33. Bayesian consensus tree resulting from the analysis of the cytb sequences from Stierandova, S. et al. 2015. Numbers in the nodes representing Bayesian posterior probability, MP, NJ and ML values. Figure adapted from Steirandova et al., 2015.

## 6.2 Discussion of cytochrome oxidase subunit I signals

European clade is poorly supported as well as the further division inside the clade. The node between *A. thessalicus*, *A. b. strymonicus* and *A. bipunctatus* has a low value either in ML tree – 0.619 either in Bayesian – 0.563 or MP tree with 53% bootstrap. Some further from Balkan region species were included to dataset (*A. fangfangae*, *A. devoli*) but they hadn't formed separate clades.

Ciscaucasian clade was supported by high values (likelihood = 98; posterior probability = 0.996; bootstrap of 98%). There were no sequences in the public databases to enrich the dataset by representatives of that clade (for example *A. kubanicus* or *A. fasciatus*).

COI is the gene that usually is used for barcoding of organisms (Hebert et al., 2003). The one of goals of this study was to provide barcoding data for spirilins from Azerbaijan and compare them with sequences from the recently published data (Jouladeh-Roudbar et al., 2016) from different localities. Sequences from eight recently described Iranian species (*A. holciki* Coad & Bogutskaya, 2012; *A. idignensis* Bogutskaya & Coad, 2009; *A. namaki* Bogutskaya & Coad 2009; *A. nicolausi* Bogutskaya & Coad, 2009; *A. qanati* Coad & Bogutskaya, 2009; *A.*

*samii* Mousavi-Sabet, Vatandoust & Doadrio, 2015; *A. tabarestanensis* Mousavi-Sabet, AnvariFar & Azizi, 2015) were added to dataset. Obtained trees have a complicated topology and species from Iran form relatively well supported clade separate from my samples from Azerbaijan. Species *A. holciki*, *A. idignensis*, *A. namaki*, *A. nicolausi*, *A. tabarestanensis* clustered within the EichS clade, while the species *A. qanati* clustered with the EichS+N clade (see e.g. Fig. 18). Based on this results, my samples do not for a monophyletic clade, and, therefore, they should not be considered as *A. eichwaldii*. Nevertheless, I decided to keep this name as a working label within this thesis, until the taxonomy and systematics of *Alburnoides* in Azerbaijan will be fully understood and clarified. The *A. holciki* has a suspicious support of 90/0.747/57 rates for ML, Bayesian PP, MP, respectively, in my analyses, while it has been fully supported by 100% of posterior probability in the original study (Roudbar et al., 2016).

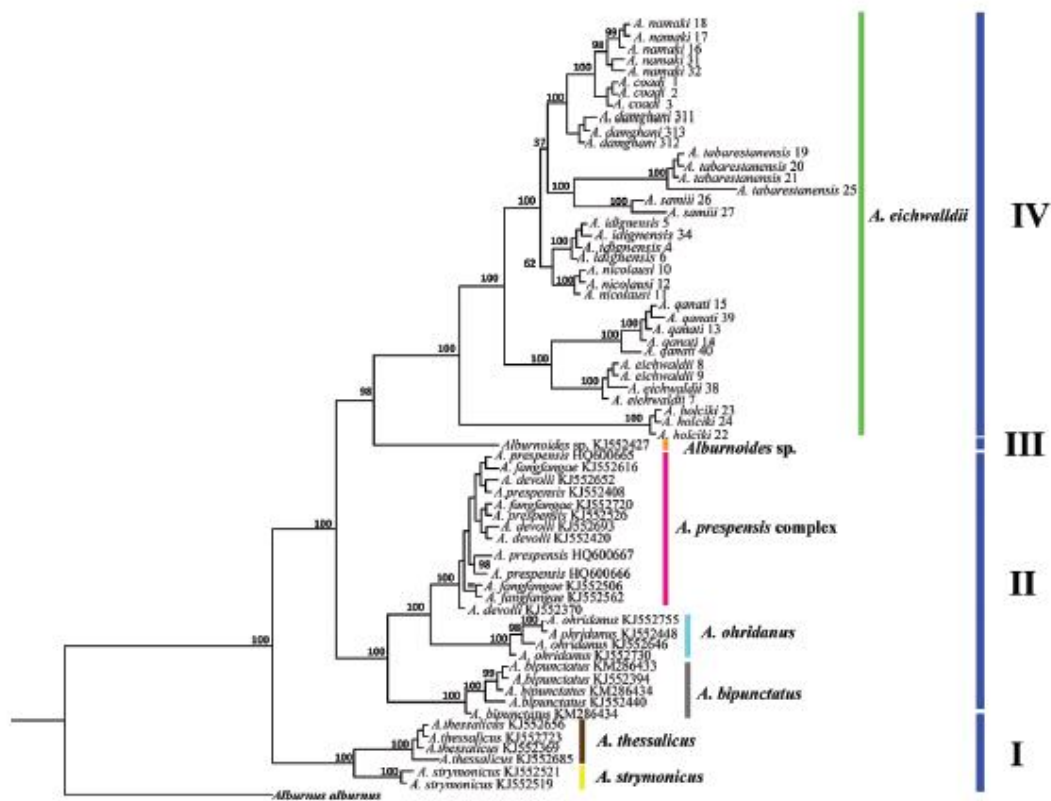


Figure 34. Bayesian tree based on analysis of COI (adapted from Roudbar et al., 2016). Sequences presented in this tree are included to dataset for analyses based on COI gene.

*Alburnoides samii* formed one clade with sequences from the Talysh mountains hydrological network (EichS). Possible explanation is that samples that were identified as *Alburnoides eichwaldii* actually belong to *A. samii*. But *A. samii* was defined as fish that is

known only from Sefidroud river that flows into the southern Caspian Sea (Roudbar et al., 2016). Obtained results raise doubts on the *A. samii* distribution range. Furthermore, validity of all de-novo described fishes is very uncertain. Genetic distances between AZ15-B-ALB-2 specimen (from Istisuchai stream, border with Iran), which together with downloaded sequences of *A. samii* forms one clade in the trees based on the COI analyses are: for cytochrome b – between AZ15-B-ALB-2 and representatives of Eur clade  $p = 7.271 - 8.166\%$ ; AZ15-B-ALB-2 and Cis clade  $p = 5.93 - 5.81\%$ ; between AZ15-B-ALB-2 and EichN+S clade  $p = 3.468 - 3.556\%$ ; between AZ15-B-ALB-2 and close related *Alburnus a. hohenackeri*  $p = 14.493\%$ . For COI gene the values are: between AZ15-B-ALB-2 and Eur clade  $p = 7.223 - 8.126\%$ ; between AZ15-B-ALB-2 and Cis clade  $p = 4.74\%$ ; between AZ15-B-ALB-2 and EichN+S clade  $p = 4.74\% - 5.869\%$ ; between AZ15-B-ALB-2 and *Alburnus a. hohenackeri* – 13.544. These are relatively high values, but they were obtained from mitochondrial genes. Nuclear RAG I obviously more conservative and has lower values: ( $p$  B15-EUR = 0.26%;  $p$  B15-Cis = 0.334;  $p$  B15-EichN+S = 0.149-0.186;  $p$  B15 – *Alburnus a. hohenackeri* = 1.374). The important point – difference between AZ15-B-ALB-2 and *A. sp 5* is 2 times bigger than distances between that specimen and representatives from EichN+S clade collected in northern localities (e.g. AZ15-L-ALB-3). P-values obtained from mitochondrial signal are around 5% both for Cis and Eich N+S clades. Hypothetically we could consider the species from EichS clade as *A. samii* but necessity of such intrageneric separation is disputable. Obviously authors provided a morphological evidence for species description, but it's unlikely that those species are reproductively isolated and have significantly different ecological features. Still, it's a question of general approach to species concept.

### 6.3 Controversy of the Eich N+S clade

The problematics of Eich N+S clade is in persistence of two inner clades that are supported by high ratios in ML, MP and Bayesian analyses e.g. from cytochrome b(100, 89%, 100). Two inner clades are formed by samples both from Greater Caucasus and Talysh Mountains hydrological network and can be easily identified inside almost all trees presented in this study except nuclear markers where presence of clades is very uncertain. The haplotype network shows that similar haplotype is shared by samples both from northern and southern localities (see Figure 31). The working hypothesis is that the fish from distant regions with different geological history would have different haplotypes. Observed results showed that *Alburnoides eichwaldii* in Azerbaijan has a mixed population structure and,

therefore, it rejected the aforementioned hypothesis of the pattern of populations explained by the simple geographical separation.

#### 6.4 Biogeographical scenario

The *Alburnoides* lineage diverged from other leuciscines cca 7.5 mya. Alpine orogeny, Paratethys formation or Eocene terminal extinction probably had no impact on modern population structure of spirilins in Azerbaijan. Those events hypothetically impacted the ancestors of cyprinid fishes (Cavender, 1991) and involved modern compound and distribution of family in the research region, but processes, which are behind the *Alburnoides* genus formation had taken place more recently.

The *Alburnoides* sp. 5 lineage diverged from the rest of the genus cca 2.2 mya (see Figure 30). The possible evolutionary scenario is discussed. Till Late Pliocene common ancestor for the *A. eichwaldii* and *A. sp.5* had a “circumparatethian” distribution widely distributed in the rivers of Ackhagilian basin (modern Caspian Sea). The salinity was around 5-8ppt (Popov, 2006) comparing to modern 12-13ppt (Ibrayev, R. A., 2010). Lower salinity probably allowed invasion and dispersion of the *Alburnoides* ancestor with spreading along and following rivers invasion. Since late Pliocene the salinity of the Caspian Sea gradually increased (Leroy, 2014), which probably limited dispersal movements. Together with final stages of Greater Caucasus formation and enclosure of the Ciscaucasian straight (see Figure 9) the populations were probably separated by the Caucasus mountain range and dispersion through brackish waters of the Caspian Sea was limited by growing salinity. Furthermore, currents of the Caspian Sea had probably changed their pattern (Hoogendoorn et al., 2005). This vicariation probably caused the *A. sp. 5* divergence from the rest of the *Alburnoides* recently distributed in the Kura and the coastal river basins of Azerbaijan. The specimens of *A. sp. 5*. are close to *A. fasciatus* and *A. kubanicus* in the phylogenetic trees. Those vouchers came from Russia, namely Sache and Abin rivers (Mendel 2010; Stierandova 2015), which are situated northwardly to greater Caucasus range and flows into Black and Azov Sea, respectively. We can assume that those species are distributed north to Caucasus in drainages that belong to paratethyan remnants (Black, Azov and Caspian Sea) and vicariation by Greater Caucasus with growing salinity therefore seems to be a plausible explanation of the observed pattern.

The Lenkaran, Vileschay and Astarachay rivers are coastal drainages in the south of Azerbaijan near the border with Iran. Samples from there were genetically heterogeneous represented by haplotypes strictly from the south, as well as haplotypes that were also found

in the northern localities contributed to Greater Caucasus hydrological network. Paraphrasing that, even the same haplotype was detected for specimens from both in southern and northern localities assuming very recent gene flow between the regions. The clade formed only by samples from the south was detected in the phylogenetic trees (e.g. Figures 16, 19, 22 – marked EichS and yellow color). This clade diverged from other sampled *Alburnoides* somewhen in-between 2.2 - 0.8mya. Hypothetically, the observed lineage richness in the southern region could be explained by persistence of the glacial refugium. Samples from the same locality descends into different clades in Bayesian, ML, MP analyses and represents different haplotypes. In detail, it is the case for locality AZ15B – river Istisuchay flowing in the Astarachay river. Three samples from that stream were used for molecular analysis. All of them took their position in different clades. This locality was the southernmost locality of our sampling area. I consider unlikely that three sympatric species of the same ecological niche would inhabit one small river. Furthermore, we were not able to distinguish the species in the field based on the morphological characters, so the existence of three independent lineage remained cryptic to me until I received the results of the genetic analysis. For confirmation of these findings and potential implications to taxonomy morphological analysis would be necessary.



Figure 35. Three specimens of *Alburnoides eichwaldii* captured in Istisuchay river (border with Iran). Each of them representing different clades.

In theory, spirilins from the north could expand to the southern regions in the interglacial periods. After ice-melting and subsequent marine transgression the spirilins could

be washed away into the sea and spread by the currents. Probably, spirilins from Greater Caucasus hydrological network have colonized rivers in Talysh mountains during the Pleistocene's transgressions.

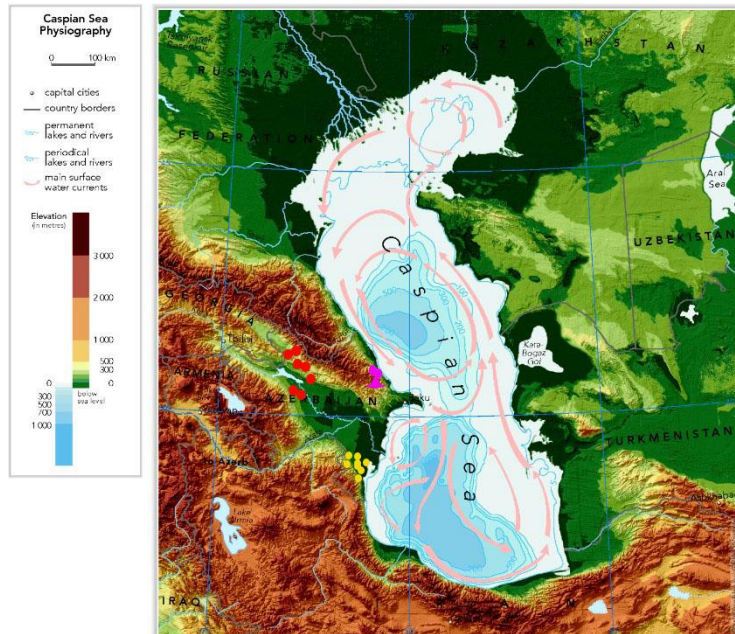


Figure 36. Currents of Caspian Sea with distribution of different *Alburnoides* lineages (red, violet and yellow dots).

### 6.5 Putatively undescribed species, *Alburnoides* sp. 5 in Ciscaucasian region

Specimens of *Alburnoides* sp 5. formed well supported clades (See Figures 15-28) in all types of analysis, except analyses based on nuclear rhodopsin genes. Likely, it can be described as new species. Further morphological analysis would be necessary.



Figure 37. *Alburnoides bipunctatus* from Ulichka, Danube drainage (Slovakia)



Figure 38. *Alburnoides* sp. 5 from Gilgichay River, Azerbaijan



Figure 39. *Alburnoides eichwaldii* from Tangaruchay river, Azerbaijan.

#### 6.6 *Hemiculter leucisculus* new invasive species for Azerbaijan fauna

During the field collections, one of the samples appeared to me like strange-looking bleak and that was the reason why I provided the molecular data for this individual. Using BLAST tool, I identified the species as a Korean sharpbelly (*Hemiculter leucisculus* Basilewsky, 1855). At the moment of identification, it was a new species report for the Azerbaijan fauna. However, at the end of 2015 there has been published a report of Korean sharpbelly in Azerbaijan freshwaters (Mustafayev et al., 2015), which confirmed my identification. It is relatively small freshwater fish native to China, Korea and Japan rivers (Berg, 1949) and it has probably been introduced later in the middle East region (reported from Iran since 1960s, Holčík and Razavi, 1992). It was barely impossible for me to identify that fish to the species level without usage of molecular methods, due to the lack of experience in work with that genus. All sequences of cytochrome b were blasted and compared to already published data and they appeared much more phylogenetically distant from spirlins and bleaks what doesn't supports idea of Alburninae sensu Howes (Howes 1991). As it was mentioned in the genus overview, Howes included *Hemiculter*, *Alburnus* and

*Alburnoides* to one subfamily. But in the phylogenetic trees specimens of *Rutilus rutilus* seems closer related than *Hemiculter*. According to the last reconsideration *Hemiculter* is within Oxygastrinae subfamily (Tang et al., 2013).



## 7 Conclusions:

1. High level of genetic variability has been found within the genus *Alburnoides* in Azerbaijan.
2. Examined individuals represent three well-supported major clades in phylogenetic trees, some of them (but not all) corresponding to the geographically isolated localities.
3. The population structure and haplotype distribution within the *Alburnoides* genus cannot be explained only by geographical isolation.
4. There is a putatively undescribed species, *Alburnoides* sp. 5 in Ciscaucasian region, a lineage which probably diverged as a result of vicariant event.
5. Identified lineage richness within the samples from the Talysh Mountains hydrological network probably bring the evidence for the existence of glacial refugium in the region.
6. The regions in the North (Greater Caucasus hydrological network) were genetically uniform, while the localities in the South (Talysh Mountains hydrological network) were composed of two different lineages. This brings an evidence for putative colonization events from the North to the South, which might have been facilitated by the known currents in the Caspian Sea.
7. Newly described invasive species for Azerbaijan fauna (*Hemiculter leucisculus*) was registered.

## 8 Literature:

- Abell, R., Thieme, M. L., Revenga, C., Bryer, M., Kottelat, M., Bogutskaya, N., ... & Stiassny, M. L. (2008). Freshwater ecoregions of the world: a new map of biogeographic units for freshwater biodiversity conservation. *BioScience*, 58(5), 403-414.
- Adamia, S., Zakariadze, G., Chkhotua, T., Sadradze, N., Tsereteli, N., Chabukiani, A., & Gventsadze, A. (2011). Geology of the Caucasus: a review. *Turkish Journal of Earth Sciences*, 20(5), 489-544.
- Akaike, H. (1976). Canonical correlation analysis of time series and the use of an information criterion. *Mathematics in Science and Engineering*, 126, 27-96.
- Almaça, C. (1976). La spéciation chez les Cyprinidae de la Péninsule Ibérique. *Revue des Travaux de l'Institut des Pêches Maritimes*, 40, 399-411.
- Azizi, F., Anvarifar, H., & Mousavi-Sabet, H. (2015). Morphological Differentiation Between Isolated Populations of Caspian spirin (*Alburnoides eichwaldii*)(Pisces: Cyprinidae) Affected by Dam.
- Banarescu, P. (1991). Zoogeography of fresh waters. Volume 2: distribution and dispersal of freshwater animals in North America and Eurasia. *Zoogeography of fresh waters. Volume 2: distribution and dispersal of freshwater animals in North America and Eurasia.*, 519-1091.
- Bănărescu, P., & Coad, B. W. (1991). Cyprinids of Eurasia. In *Cyprinid Fishes*(pp. 127-155). Springer Netherlands.
- Bandelt, H. J., Forster, P., & Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular biology and evolution*, 16(1), 37-48.
- Berg, L. S. (1940). Zoogeografiya presnovodnykh ryb Perednei Azii (Zoogeography of freshwater fish of the Near East). *Uchenye Zapiski leningradskogo gosudarstvennogo Universiteta Seriya Geograficheskikh Nauk*,3, 3-31.
- Berg, L. S. (1949). Presnovodnye ryby Irana i sopredel'nykh stran [Freshwater fishes of Iran and adjacent countries]. *Trudy Zoologicheskogo Instituta Akademii Nauk SSSR*, 8, 783-858.
- Bianco, P. G. (1990). Potential role of the palaeohistory of the Mediterranean and Paratethys basins on the early dispersal of Euro-Mediterranean freshwater fishes. *Ichthyological exploration of freshwaters. Munchen*, 1(2), 167-184.
- Bogutskaya, N. G., & Coad, B. W. (2009). A review of vertebral and fin-ray counts in the genus *Alburnoides* (Teleostei: Cyprinidae) with a description of six new species. *Zoosystematica Rossica*, 18(1), 126-173.
- Bogutskaya, N. G., & Naseka, A. M. (2004). Katalog beschelyustnykh i ryb presnykh i solonovatykh vod Rossii s nomenklaturnymi i taksonomicheskimi kommentariyami. Zoological Institute, Russian Academy of Sciences and KMK Scientific Press Ltd, Moscow.

- Bogutskaya, N. G., Zupancic, P., & Naseka, A. M. (2010). Two new species of freshwater fishes of the genus *Alburnoides*, *A. fangfangae* and *A. devolli* (Actinopterygii: Cyprinidae), from the Adriatic Sea basin in Albania. *Proceedings of the Zoological Institute*, 314(4), 448-468.
- Bohlen, J., Perdices, A., Doadrio, I., & Economidis, P. S. (2006). Vicariance, colonisation, and fast local speciation in Asia Minor and the Balkans as revealed from the phylogeny of spined loaches (Osteichthyes; Cobitidae). *Molecular Phylogenetics and Evolution*, 39(2), 552-561.
- Briolay, J., Galtier, N., Brito, R. M., & Bouvet, Y. (1998). Molecular phylogeny of cyprinidae inferred from cytochrome bDNA Sequences. *Molecular phylogenetics and evolution*, 9(1), 100-108.
- Bruto, R. M., Briolay, J., Galtier, N., Bouvet, Y., & Coelho, M. M. (1997). Phylogenetic Relationships within Genus *Leuciscus* (Pisces, Cyprinidae) in Portuguese Fresh Waters, Based on Mitochondrial DNA Cytochrome b Sequences. *Molecular Phylogenetics and Evolution*, 8(3), 435-442.
- Cavender, T. M. (1991). The fossil record of the Cyprinidae. In *Cyprinid fishes* (pp. 34-54). Springer Netherlands.
- Cavender, T. M., & Coburn, M. M. (1992). Phylogenetic relationships of North American cyprinidae. *Systematics, historical ecology, and North American freshwater fishes*, 293-327.
- Chen, X.L., Yue, P.Q., and Lin R.D. (1984). Major groups within the family Cyprinidae and their phylogenetic relationships. *Acta Zootaxonomica Sinica* 9: 424-440.
- Chen, W. J., Bonillo, C., & Lecointre, G. (2003). Repeatability of clades as a criterion of reliability: a case study for molecular phylogeny of Acanthomorpha (Teleostei) with larger number of taxa. *Molecular phylogenetics and evolution*, 26(2), 262-288.
- Clauzon, G., Suc, J. P., Gautier, F., Berger, A., & Loutre, M. F. (1996). Alternate interpretation of the Messinian salinity crisis: Controversy resolved?. *Geology*, 24(4), 363-366.
- Coad, B. W., & Bogutskaya, N. G. (2012). A new species of riffle minnow, *Alburnoides holciki*, from the Hari River basin in Afghanistan and Iran (Actinopterygii: Cyprinidae). *Zootaxa*, 3453, 43-55.
- Cox, C. B. (2001). The biogeographic regions reconsidered. *Journal of Biogeography*, 28(4), 511-523.
- Croizat, L., Nelson, G., & Rosen, D. E. (1974). Centers of origin and related concepts. *Systematic Biology*, 23(2), 265-287.
- Cunha, C., Mesquita, N., Dowling, T. E., Gilles, A., & Coelho, M. M. (2002). Phylogenetic relationships of Eurasian and American cyprinids using cytochrome b sequences. *Journal of Fish Biology*, 61(4), 929-944.

- Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nature methods*, 9(8), 772-772.
- Drummond, A. J., Ho, S. Y., Phillips, M. J., & Rambaut, A. (2006). Relaxed phylogenetics and dating with confidence. *PLoS Biol*, 4(5), e88.
- Drummond, A. J., Suchard, M. A., Xie, D., & Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular biology and evolution*, 29(8), 1969-1973.
- Durand, J. D., Tsigenopoulos, C. S., Ünlü, E., & Berrebi, P. (2002). Phylogeny and biogeography of the family Cyprinidae in the Middle East inferred from cytochrome b DNA—evolutionary significance of this region. *Molecular Phylogenetics and Evolution*, 22(1), 91-100.
- Esmaili, H. R., Coad, B. W., Gholamifard, A., Nazari, N., & Teimory, A. (2010). Annotated checklist of the freshwater fishes of Iran. *Zoosystematica Rossica*, 19(2), 361-386.
- Eschmeyer, W. N. (2015). Catalog of Fishes: genera, species, references (<http://research.calacademy.org/research/ichthyology/catalog/fishcatmain.asp>). Accessed Jun. 2016
- Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, 783-791.
- Geiger, M. F., Herder, F., Monaghan, M. T., Almada, V., Barbieri, R., Bariche, M., ... & Denys, G. P. (2014). Spatial heterogeneity in the Mediterranean Biodiversity Hotspot affects barcoding accuracy of its freshwater fishes. *Molecular ecology resources*, 14(6), 1210-1221.
- Gilles, A., Lecointre, G., Faure, E., Chappaz, R., & Brun, G. (1998). Mitochondrial phylogeny of the European cyprinids: implications for their systematics, reticulate evolution, and colonization time. *Molecular Phylogenetics and Evolution*, 10(1), 132-143.
- Gilles, A., Lecointre, G., Miquelis, A., Loerstcher, M., Chappaz, R., & Brun, G. (2001). Partial combination applied to phylogeny of European cyprinids using the mitochondrial control region. *Molecular Phylogenetics and Evolution*, 19(1), 22-33.
- Guindon, S., & Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic biology*, 52(5), 696-704.
- Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic biology*, 59(3), 307-321.
- He, S., Mayden, R. L., Wang, X., Wang, W., Tang, K. L., Chen, W. J., & Chen, Y. (2008). Molecular phylogenetics of the family Cyprinidae (Actinopterygii: Cypriniformes) as evidenced by sequence variation in the first intron of S7 ribosomal protein-coding gene: Further evidence from a nuclear gene of the systematic chaos in the family. *Molecular phylogenetics and evolution*, 46(3), 818-829.
- Hebert, P. D., Cywinska, A., & Ball, S. L. (2003). Biological identifications through DNA

- barcodes. *Proceedings of the Royal Society of London B: Biological Sciences*, 270(1512), 313-321.
- Hewitt, G. M. (1999). Post-glacial re-colonization of European biota. *Biological journal of the Linnean Society*, 68(1-2), 87-112.
- Holcik, J., & Razavi, B. A. (1992). On some new or little known fresh-water fishes from the Iranian coast of the Caspian sea.
- Howes, G. J. (1991). Systematics and biogeography: an overview. In *Cyprinid fishes* (pp. 1-33). Springer Netherlands.
- Hoogendoorn, R. M., Boels, J. F., Kroonenberg, S. B., Simmons, M. D., Aliyeva, E., Babazadeh, A. D., & Huseynov, D. (2005). Development of the Kura delta, Azerbaijan; a record of Holocene Caspian sea-level changes. *Marine Geology*, 222, 359-380.
- Huelsenbeck, J. P., & Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, 17(8), 754-755.
- Ibrayev, R. A., Özsoy, E., Schrum, C., & Sur, H. I. (2010). Seasonal variability of the Caspian Sea three-dimensional circulation, sea level and air-sea interaction. *Ocean Science*, 6(1).
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Mentjies, P., & Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, 28(12), 1647-1649.
- Kingman, J. F. C. (1982). The coalescent. *Stochastic processes and their applications*, 13(3), 235-248.
- Kottelat, M., & Freyhof, J. (2007). *Handbook of European freshwater fishes*. Publications Kottelat
- Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular biology and evolution*, msw054.
- Machordom A, Doadrio I. (2001). Evidence of a Cenozoic Betic-Kabilian connection based on freshwater fish phylogeography (*Luciobarbus*, Cyprinidae). *Mol Phylogenet Evol* 18(2): 252-263.
- Mamedov, A. V. (1997). The late Pleistocene-Holocene history of the Caspian sea. *Quaternary International*, 41, 161-166.
- Mau, B., & Newton, M. A. (1997). Phylogenetic inference for binary data on dendrograms using Markov chain Monte Carlo. *Journal of Computational and Graphical Statistics*, 6(1), 122-131.

- Marková, S., Šanda, R., Crivelli, A., et al., 2010. Nuclear and mitochondrial DNA sequence data reveal the evolutionary history of *Barbus* (Cyprinidae) in the ancient lake systems of the Balkans. *Mol. Phylogenet. Evol.* 55, 488–500.
- Mousavi-Sabet, H., Vatandoust, S., & Doadrio, I. (2015). Review of the genus *Alburnoides* Jeitteles, 1861 (Actinopterygii, Cyprinidae) from Iran with description of three new species from the Caspian Sea and Kavir basins.
- Mustafayev, N. J., Ibrahimov, S. R., & Levin, B. A. (2015). Korean sharpbelly *Hemiculter leuciscus* (Basilewsky, 1855)(Cypriniformes, Cyprinidae) is a new species of Azerbaijan fauna. *Russian Journal of Biological Invasions*, 6(4), 252-259.
- Myers, N. (1988). Threatened biotas: "hot spots" in tropical forests. *Environmentalist*, 8(3), 187-208.
- Naseka, A. M., & Bogutskaya, N. G. (2009). Fishes of the Caspian Sea: zoogeography and updated check-list. *Zoosystematica Rossica*, 18(2), 295-317.
- Naseka, A. M. (2010). Zoogeographical freshwater divisions of the Caucasus as a part of the West Asian Transitional Region. *Proceedings of the Zoological Institute, Russian Academy of Sciences*, 314(4), 469-492.
- Nei, M., & Kumar, S. (2000). *Molecular evolution and phylogenetics*. Oxford university press.
- Nelson, J. S. (2006). *Fishes of the world*. New York: J.
- Perdices, A., Doadrio, I., Economidis, P. S., Bohlen, J., & Bănărescu, P. (2003). Pleistocene effects on the European freshwater fish fauna: double origin of the cobitid genus *Sabanejewia* in the Danube basin (Osteichthyes: Cobitidae). *Molecular Phylogenetics and Evolution*, 26(2), 289-299.
- Perea, S., Böhme, M., Zupančič, P., Freyhof, J., Šanda, R., Özuluğ, M., ... & Doadrio, I. (2010). Phylogenetic relationships and biogeographical patterns in Circum-Mediterranean subfamily Leuciscinae (Teleostei, Cyprinidae) inferred from both mitochondrial and nuclear data. *BMC Evolutionary Biology*, 10(1), 1
- Popov, S. V., Rögl, F., Rozanov, A. Y., Steininger, F. F., Shcherba, I. G., & Kovac, M. (2004). Lithological-Paleogeographic maps of Paratethys-10 maps Late Eocene to Pliocene.
- Popov, S. V., Shcherba, I. G., Ilyina, L. B., Nevesskaya, L. A., Paramonova, N. P., Khondkarian, S. O., & Magyar, I. (2006). Late Miocene to Pliocene palaeogeography of the Paratethys and its relation to the Mediterranean. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 238(1), 91-106
- Prothero, D. R. (1989). Stepwise extinctions and climatic decline during the later Eocene and Oligocene. *Mass extinctions: Processes and evidence*, 217-234.
- Reyjol, Y., Hugueny, B., Pont, D., Bianco, P. G., Beier, U., Caiola, N., ... & Haidvogel, G.

(2007). Patterns in species richness and endemism of European freshwater fish. *Global Ecology and Biogeography*, 16(1), 65-75.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., ... & Huelsenbeck, J. P. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic biology*, 61(3), 539-542.

Roudbar, A. J., Eagderi, S., Esmaeili, H. R., Coad, B. W., & Bogutskaya, N. (2016). A molecular approach to the genus *Alburnoides* using COI sequences data set and the description of a new species, *A. damghani*, from the Damghan River system (the Dasht-e Kavir Basin, Iran) (Actinopterygii, Cyprinidae). *ZooKeys*, (579), 157.

Quenouille, B., Bermingham, E., & Planes, S. (2004). Molecular systematics of the damselfishes (Teleostei: Pomacentridae): Bayesian phylogenetic analyses of mitochondrial and nuclear DNA sequences. *Molecular phylogenetics and evolution*, 31(1), 66-88.

Rögl, F. (1999). Mediterranean and Paratethys. Facts and hypotheses of an Oligocene to Miocene paleogeography (short overview). *Geologica carpathica*, 50(4), 339-349.

Seifali, M., Arshad, A., Moghaddam, F. Y., Esmaeili, H. R., Kiabi, B. H., Daud, S. K., & Aliabadian, M. (2012). Mitochondrial genetic differentiation of spirin (Actinopterygii: Cyprinidae) in the south Caspian Sea basin of Iran. *Evolutionary bioinformatics online*, 8, 219.

Smith, K. G., Barrios, V., Darwall, W. R., & Numa, C. (Eds.). (2014). The status and distribution of freshwater biodiversity in the Eastern Mediterranean. IUCN.

Stierandová, S., Vukić, J., Vasil'eva, E. D., Zogaris, S., Shumka, S., Halačka, K., ... & Koščo, J. (2015). A multilocus assessment of nuclear and mitochondrial sequence data elucidates phylogenetic relationships among European spirins (*Alburnoides*, Cyprinidae). *Molecular phylogenetics and evolution*, 94, 479-491.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., & Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular biology and evolution*, 28(10), 2731-2739.

Tang, K. L., Agnew, M. K., Hirt, M. V., Lumbantobing, D. N., Raley, M. E., Sado, T., ... & He, S. (2013). Limits and phylogenetic relationships of East Asian fishes in the subfamily Oxygastrinae (Teleostei: Cypriniformes: Cyprinidae). *Zootaxa*, 3681(2), 101-135.

Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic acids research*, 22(22), 4673-4680..

- Turan, D., Ekmekçi, F. G., Kaya, C., & Güçlü, S. S. (2013). *Alburnoides manyasensis* (Actinopterygii, Cyprinidae), a new species of cyprinid fish from Manyas Lake basin, Turkey. *ZooKeys*, (276), 85.
- Turan, D., Kaya, C. Ü. N. E. Y. T., Ekmekçi, F. G., & Doğan, E. S. R. A. (2014). Three new species of *Alburnoides* (Teleostei: Cyprinidae) from Euphrates River, Eastern Anatolia, Turkey. *Zootaxa*, 3754(2), 101-116.
- Turan, D., BEKTAŞ, Y., KAYA, C., & Baycelebi, E. (2016). *Alburnoides diclensis* (Actinopterygii: Cyprinidae), a new species of cyprinid fish from the upper Tigris River, Turkey. *Zootaxa*, 4067(1), 79-87.
- Wang, X., Gan, X., Li, J., Mayden, R. L., & He, S. (2012). Cyprinid phylogeny based on Bayesian and maximum likelihood analyses of partitioned data: implications for Cyprinidae systematics. *Science China Life Sciences*, 55(9), 761-773.
- Ward, R. D., Zemlak, T. S., Innes, B. H., Last, P. R., & Hebert, P. D. (2005). DNA barcoding Australia's fish species. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 360(1462), 1847-1857.
- Winfield, I., & Nelson, J. S. (Eds.). (2012). *Cyprinid fishes: systematics, biology and exploitation* (Vol. 3). Springer Science & Business Media.
- Yang, Z., & Rannala, B. (1997). Bayesian phylogenetic inference using DNA sequences: a Markov Chain Monte Carlo method. *Molecular biology and evolution*, 14(7), 717-724.
- Zardoya, R., & Doadrio, I. (1999). Molecular evidence on the evolutionary and biogeographical patterns of European cyprinids. *Journal of Molecular Evolution*, 49(2), 227-237.
- Zar, K. (1982). *National Geographic Atlas of the World*. The Library, 52(3).



## 9 Appendix

**Appendix 1.** Table: detailed review of specimens used for molecular analysis.

Sample code	Species	Sampling site	Latitude	Longitude	Basin/drainage
SK15-B-ALB-2	<i>Alburnoides bipunctatus</i>	Ulichka river	48°56'53.59"	22°26'17.26"	Danube River
SK15-B-ALB-1	<i>Alburnoides bipunctatus</i>	Ulichka river	48°56'53.59"	22°26'17.26"	Danube River
SK15-A-ALB-3	<i>Alburnoides bipunctatus</i>	Udava river	48°58'06.90"	21°57'2990"	Danube River
SK15-A-ALB-2	<i>Alburnoides bipunctatus</i>	Udava river	48°58'06.90"	21°57'2990"	Danube River
SK15-A-ALB-1	<i>Alburnoides bipunctatus</i>	Udava river	48°58'06.90"	21°57'2990"	Danube River
AZ15-O-ALB-3	<i>Barbus lacerta</i>	unnamed flood	41°39'47.31"	46°34'2.50"	Katehchay River
AZ15-O-ALB-2	<i>Alburnoides eichwaldii</i>	unnamed flood	41°39'47.31"	46°34'2.50"	Katehchay River
AZ15-O-ALB-1	<i>Alburnoides eichwaldii</i>	unnamed flood	41°39'47.31"	46°34'2.50"	Katehchay River
AZ15-M-ALB-4	<i>Alburnoides eichwaldii</i>	Balakanchay River	41°43'36.91"	46°27'28.72"	Alazani River
AZ15-M-ALB-3	<i>Alburnoides eichwaldii</i>	Balakanchay River	41°43'36.91"	46°27'28.72"	Alazani River
AZ15-M-ALB-2	<i>Alburnoides eichwaldii</i>	Balakanchay River	41°43'36.91"	46°27'28.72"	Alazani River
AZ15-M-ALB-1	<i>Alburnoides eichwaldii</i>	Balakanchay River	41°43'36.91"	46°27'28.72"	Alazani River
AZ15-L-ALB-5	<i>Alburnus a. hohenackeri</i>	Qarachay River	41°28'25.56"	46°27'17.02"	Alazani River
AZ15-L-ALB-4	<i>Alburnus a. hohenackeri</i>	Qarachay River	41°28'25.56"	46°27'17.02"	Alazani River
AZ15-L-ALB-3	<i>Alburnoides eichwaldii</i>	Qarachay River	41°28'25.56"	46°27'17.02"	Alazani River
AZ15-L-ALB-2	<i>Alburnoides eichwaldii</i>	Qarachay River	41°28'25.56"	46°27'17.02"	Alazani River
AZ15-L-ALB-1	<i>Alburnoides eichwaldii</i>	Qarachay River	41°28'25.56"	46°27'17.02"	Alazani River
AZ15-K-ALB-3	<i>Alburnoides eichwaldii</i>	Katehchay River	41°42'45.57"	46°36'03.95"	Alazani River
AZ15-K-ALB-2	<i>Alburnoides eichwaldii</i>	Katehchay River	41°42'45.57"	46°36'03.95"	Alazani River
AZ15-K-ALB-1	<i>Alburnoides eichwaldii</i>	Katehchay River	41°42'45.57"	46°36'03.95"	Alazani River
AZ15-J-ALB-3	<i>Alburnoides eichwaldii</i>	Limanchay River	41°27'26.90"	46°40'32.63"	Alazani River
AZ15-J-ALB-2	<i>Alburnoides eichwaldii</i>	Limanchay River	41°27'26.90"	46°40'32.63"	Alazani River
AZ15-J-ALB-1	<i>Alburnoides eichwaldii</i>	Limanchay River	41°27'26.90"	46°40'32.63"	Alazani River
AZ15-I-ALB-1	<i>Alburnoides eichwaldii</i>	Agrichay River	41°14'24.85"	46°54'25.73"	Alazani River
AZ15-F-ALB-3	<i>Alburnoides eichwaldii</i>	Lankaran River	38°42'42.62"	48°44'59.52"	Caspian Sea

AZ15-F-ALB-2	<i>Alburnoides eichwaldii</i>	Lankaran River	38°42'42.62"	48°44'59.52"	Caspian Sea
AZ15-F-ALB-1	<i>Alburnoides eichwaldii</i>	Lankaran River	38°42'42.62"	48°44'59.52"	Caspian Sea
AZ15-E-ALB-3	<i>Alburnoides eichwaldii</i>	Lankaran River	38°44'10.30"	48°50'17.91"	Caspian Sea
AZ15-E-ALB-2	<i>Alburnoides eichwaldii</i>	Lankaran River	38°44'10.30"	48°50'17.91"	Caspian Sea
AZ15-E-ALB-1	<i>Alburnoides eichwaldii</i>	Lankaran River	38°44'10.30"	48°50'17.91"	Caspian Sea
AZ15-CH-ALB-3	<i>Alburnoides eichwaldii</i>	Katehchay River	41°40'24.06"	46°34'15.93"	Alazani River
AZ15-CH-ALB-2	<i>Alburnoides eichwaldii</i>	Katehchay River	41°40'24.06"	46°34'15.93"	Alazani River
AZ15-CH-ALB-1	<i>Alburnoides eichwaldii</i>	Katehchay River	41°40'24.06"	46°34'15.93"	Alazani River
AZ15-C-ALB-3	<i>Alburnoides eichwaldii</i>	Lankaran River	38°44'7.02"	48°37'51.75"	Caspian Sea
AZ15-C-ALB-2	<i>Alburnoides eichwaldii</i>	Lankaran River	38°44'7.02"	48°37'51.75"	Caspian Sea
AZ15-C-ALB-1	<i>Alburnoides eichwaldii</i>	Lankaran River	38°44'7.02"	48°37'51.75"	Caspian Sea
AZ15-B-ALB-3	<i>Alburnoides eichwaldii</i>	Istisuchay River	38°27'01.06"	48°45'57.20"	Astarachay River
AZ15-B-ALB-2	<i>Alburnoides eichwaldii</i>	Istisuchay River	38°27'01.06"	48°45'57.20"	Astarachay River
AZ15-B-ALB-1	<i>Alburnoides eichwaldii</i>	Istisuchay River	38°27'01.06"	48°45'57.20"	Astarachay River
AZ15-A-ALB-3	<i>Alburnoides eichwaldii</i>	Tangaruchay River	38°31'28.47"	48°41'60"	Tangaruchay River
AZ15-A-ALB-2	<i>Alburnoides eichwaldii</i>	Tangaruchay River	38°31'28.47"	48°41'60"	Tangaruchay River
AZ15-A-ALB-1	<i>Alburnoides eichwaldii</i>	Tangaruchay River	38°31'28.47"	48°41'60"	Tangaruchay River
AZ14-Q-RUT-1	<i>Rutilus rutilus</i>	unnamed flood	39°40'15.35"	48°10'58.41"	Caspian Sea
AZ14-P-ALB-2	<i>Alburnoides sp</i> 5.	Gilgichay river	41°09'19.85"	49°04'14.60"	Caspian Sea
AZ14-P-ALB-1	<i>Alburnoides sp</i> 5.	Gilgichay river	41°09'19.85"	49°04'14.60"	Caspian Sea
AZ14-K-ALB-3	<i>Alburnoides sp</i> 5.	Caqucuchay River	41°15'03.22"	48°35'05.19"	Caspian Sea
AZ14-K-ALB-2	<i>Alburnoides sp</i> 5.	Caqucuchay River	41°15'03.22"	48°35'05.19"	Caspian Sea
AZ14-K-ALB-1	<i>Alburnoides sp</i> 5.	Caqucuchay River	41°15'03.22"	48°35'05.19"	Caspian Sea
AZ14-G-ALB-3	<i>Alburnoides eichwaldii</i>	Verevulchay River	38°48'39.78"	48°41'13.98"	Caspian Sea
AZ14-G-ALB-2	<i>Alburnoides eichwaldii</i>	Verevulchay River	38°48'39.78"	48°41'13.98"	Caspian Sea
AZ14-G-ALB-1	<i>Alburnoides eichwaldii</i>	Verevulchay River	38°48'39.78"	48°41'13.98"	Caspian Sea
AZ14-CH-SHE-1	<i>Alburnus filipii</i>	Lenkaran river	38°44'24.78"	48°50'31.7"	Caspian Sea
AZ14-CH-LEU-1	<i>Rutilus frisii</i>	Lenkaran river	38°44'24.78"	48°50'31.7"	Caspian Sea
AZ14-CH-ALB-3	<i>Alburnoides eichwaldii</i>	Lenkaran river	38°44'24.78"	48°50'31.7"	Caspian Sea
AZ14-CH-ALB-2	<i>Alburnus a. hohenackeri</i>	Lenkaran river	38°44'24.78"	48°50'31.7"	Caspian Sea

AZ14-CH-ALB-1	<i>Alburnus a. hohenackeri</i>	Lenkaran river	38°44'24.78"	48°50'31.7"	Caspian Sea
AZ14-C-ALB-4	<i>Alburnoides eichwaldii</i>	Vileschay River	38°58'47.48"	48°33'55"	Caspian Sea
AZ14-C-ALB-3	<i>Hemiculter leucisculus</i>	Vileschay River	38°58'47.48"	48°33'55"	Caspian Sea
AZ14-C-ALB-2	<i>Hemiculter leucisculus</i>	Vileschay River	38°58'47.48"	48°33'55"	Caspian Sea
AZ14-C-ALB-1	<i>Hemiculter leucisculus</i>	Vileschay River	38°58'47.48"	48°33'55"	Caspian Sea
AZ13-I-ALB-1	<i>Alburnoides sp</i> 5.	Atahchay River	41°03'20.88"	49°07'49.14"	Caspian Sea
AZ13-H-ALB-4	<i>Rutilus rutilus</i>	unnamed flood	41°12'26.82	49°08'13.31"	Caspian Sea
AZ13-H-ALB-3	<i>Rutilus frisii</i>	unnamed flood	41°12'26.81	49°08'13.31"	Caspian Sea
AZ13-H-ALB-2	<i>Alburnus filipii</i>	unnamed flood	41°12'26.80	49°08'13.31"	Caspian Sea
AZ13-H-ALB-1	<i>Alburnus alburnus hohenackeri</i>	unnamed flood	41°12'26.79	49°08'13.31"	Caspian Sea
AZ13-F-SQU-2	<i>Squalius cephalus</i>	Geogchay River	40°43'26.28"	47°47'20.68"	Kura River
AZ13-F-SQU-1	<i>Squalius cephalus</i>	Geogchay River	40°43'26.28"	47°47'20.68"	Kura River
AZ13-F-ALB-3	<i>Alburnoides eichwaldii</i>	Geogchay River	40°43'26.28"	47°47'20.68"	Kura River
AZ13-F-ALB-2	<i>Alburnoides eichwaldii</i>	Geogchay River	40°43'26.28"	47°47'20.68"	Kura River
AZ13-F-ALB-1	<i>Alburnoides eichwaldii</i>	Geogchay River	40°43'26.28"	47°47'20.68"	Kura River
AZ13-D-SQU-2	<i>Squalius cephalus</i>	Sululchay River	40°40'50.23"	48°29'57.72"	Kura River
AZ13-D-SQU-1	<i>Squalius cephalus</i>	Sululchay River	40°40'50.23"	48°29'57.72"	Kura River
AZ12-J-ALB-3	<i>Alburnus alburnus hohenackeri</i>	irrigation chanel	40°38'47.84"	46°40'6.95"	Kura River
AZ12-J-ALB-2	<i>Alburnus a. hohenackeri</i>	irrigation chanel	40°38'47.84"	46°40'6.95"	Kura River
AZ12-J-ALB-1	<i>Alburnus a. hohenackeri</i>	irrigation chanel	40°38'47.84"	46°40'6.95"	Kura River
AZ12-I-ALB-1	<i>Alburnoides eichwaldii</i>	Kurakchay River	40°39'47.80"	46°38'19.33"	Kura River
AZ12-H-ALB-1	<i>Squalius cephalus</i>	Turyanchay River	40°42'49.65"	47°32'56.65"	Kura River
AZ12-C-ALB-3	<i>Alburnoides sp</i> 5.	Tugchay River	40°51'55.07"	49°13'8.85"	Caspian Sea
AZ12-C-ALB-2	<i>Alburnoides sp</i> 5.	Tugchay River	40°51'55.07"	49°13'8.85"	Caspian Sea
AZ12-C-ALB-1	<i>Alburnoides sp</i> 5.	Tugchay River	40°51'55.07"	49°13'8.85"	Caspian Sea
AZ12-A-ALB-3	<i>Alburnoides sp</i> 5.	Tugchay River	40°52'43.54"	49°11'17.17"	Caspian Sea
AZ12-A-ALB-2	<i>Alburnoides sp</i> 5.	Tugchay River	40°52'43.54"	49°11'17.17"	Caspian Sea
AZ12-A-ALB-1	<i>Alburnoides sp</i> 5.	Tugchay River	40°52'43.54"	49°11'17.17"	Caspian Sea

## **Appendix 2. List of valid *Alburnoides* species:**

*Alburnoides bipunctatus* Bloch, 1782

*A. coadi* Mousavi-Sabet, Vatandoust & Doadrio, 2015

*A. damghani* Jouladeh-Roudbar, Eagderi, Esmaeili, Coad & Bogutskaya, 2016

*A. devolli* Bogutskaya, Zupančič & Naseka, 2010

*A. diclensis* Turan, Bektaş, Kaya & Bayçelebi, 2016

*A. eichwaldii* De Filippi, 1863

*A. emineae* Turan, Kaya, Ekmekçi & Doğan, 2014

*A. fasciatus* Nordmann, 1840

*A. gmelini* Bogutskaya & Coad, 2009

*A. holciki* Coad & Bogutskaya, 2012

*A. idignensis* Bogutskaya & Coad, 2009

*A. kubanicus* Bănărescu, 1964

*A. maculatus* Kessler, 1859

*A. manyasensis* Turan, Ekmekçi, Kaya & Güçlü, 2013

*A. namaki* Bogutskaya & Coad, 2009

*A. nicolausi* Bogutskaya & Coad, 2009

*A. oblongus* Bulgakov, 1923

*A. parhami* Mousavi-Sabet, Vatandoust & Doadrio, 2015

*A. petrubanarescui* Bogutskaya & Coad, 2009

*A. ohridanus* Karaman, 1928

*A. prespensis* Karaman, 1924

*A. qanati* Coad & Bogutskaya, 2009

*A. recepi* Turan, Kaya, Ekmekçi & Doğan, 2014

*A. rossicus* Berg, 1924

*A. samiii* Mousavi-Sabet, Vatandoust & Doadrio, 2015

*A. smyrnae* Pellegrin, 1927

*A. strymonicus* Chichkoff, 1940

*A. taeniatus* Kessler, 1874

*A. tabarestanensis* Mousavi-Sabet, AnvariFar & Azizi, 2015,

*A. thessalicus* Stephanidis, 1950

*A. tzanevi* Chichkoff, 1933.

*A. varentsovi* Bogutskaya & Coad, 2009

*A. velioglui* Turan, Kaya, Ekmekçi & Doğan, 2014

### Appendix 3. Catalogue of downloaded sequences:

#### Appendix 3.1 Table of downloaded cytochrome b sequences:

Species	Accession number	gene	Authors	Country
<i>A. bipunctatus</i>	HM560059	cytb	Perea,S.	France
<i>A. bipunctatus</i>	HM173131	cytb	Mendel,J.,	Czech Republic
<i>A. bipunctatus</i>	HM173111	cytb	Mendel,J.,	Czech Republic
<i>A. bipunctatus rossicus</i>	HM173134	cytb	Mendel,J.,	Russia
<i>A. bipunctatus rossicus</i>	KM874611	cytb	Stierandova,S.,	Ukraine
<i>A. bipunctatus rossicus</i>	KM874608	cytb	Stierandova,S.,	Russia
<i>A. tzanevi</i>	HM173132	cytb	Mendel,J.	Bulgary
<i>A. tzanevi</i>	KM874640	cytb	Stierandova,S.,	Bulgary
<i>A. tzanevi</i>	KM874639	cytb	Stierandova,S.,	Bulgary
<i>A. sp 2</i>	KM874505	cytb	Stierandova,S.,	Romania
<i>A. sp 2</i>	HM173148	cytb	Mendel,J.	Serbia
<i>A. sp 2</i>	HQ658899	cytb	Seifali,M	Iran
<i>A. sp 1</i>	HQ658866	cytb	Seifali,M	Iran
<i>A. sp 1</i>	HQ658865	cytb	Seifali,M	Iran
<i>A. sp 1</i>	KM874484	cytb	Stierandova,S.,	Bosnia and Herzegovina
<i>A. sp 1</i>	HM173142	cytb	Mendel,J.	Croatia
<i>A. ohridanus</i>	HM173156	cytb	Mendel,J.	Albania
<i>A. ohridanus</i>	KM874596	cytb	Stierandova,S.,	Albania
<i>A. ohridanus</i>	KM874593	cytb	Stierandova,S.,	Albania
<i>A. cf prespensis</i>	KM874601	cytb	Stierandova,S.,	Albania
<i>A. cf prespensis</i>	KM874576	cytb	Stierandova,S.,	Albania
<i>A. cf prespensis</i>	HM173161	cytb	Mendel,J.	Albania
<i>A. sp 3</i>	HM173152	cytb	Mendel,J.	Ukraine
<i>A. sp 3</i>	KM874621	cytb	Stierandova,S.,	Greece
<i>A. sp 3</i>	KM874633	cytb	Stierandova,S.,	Greece
<i>A. fasciatus</i>	HM173170	cytb	Mendel,J.	Russia
<i>A. fasciatus</i>	HM173168	cytb	Mendel,J.	Russia
<i>A. fasciatus</i>	KM874580	cytb	Stierandova,S.,	Russia
<i>A. kubanicus</i>	HM173174	cytb	Mendel,J.	Russia
<i>A. kubanicus</i>	KM874584	cytb	Stierandova,S.,	Russia
<i>A. kubanicus</i>	KM874583	cytb	Stierandova,S.,	Russia
<i>A. sp 5</i>	KM874520	cytb	Stierandova,S.,	Azerbaijan
<i>A. sp 5</i>	KM874519	cytb	Stierandova,S.,	Azerbaijan
<i>A. sp 5</i>	KM874518	cytb	Stierandova,S.,	Azerbaijan
<i>A. eichwaldii</i>	KM874573	cytb	Stierandova,S.,	Azerbaijan
<i>A. eichwaldii</i>	KM874572	cytb	Stierandova,S.,	Azerbaijan
<i>A. eichwaldii</i>	HQ658883	cytb	Seifali,M	Iran
<i>A. thessalicus</i>	HM173165	cytb	Mendel,J.	Macedonia
<i>A. thessalicus</i>	HM173164	cytb	Mendel,J.	Macedonia
<i>A. b.strymonicus</i>	HM173167	cytb	Mendel,J.	Greece
<i>A. b.strymonicus</i>	KM874618	cytb	Stierandova,S.,	Greece
<i>A. b.strymonicus</i>	KM874617	cytb	Stierandova,S.,	Greece

### Appendix 3.2 Table of downloaded cytochrome oxidase subunit 1 sequences:

Species	Accession number	gene	Authors	country
<i>A. bipunctatus</i>	HQ960561	coI	Šanda, R.	Czech Republic
<i>A. bipunctatus</i>	HM560238	coI	Perea, S.	France
<i>A. bipunctatus</i>	KM286433	coI	Knebelsberger, T.	Germany
<i>A. sp HRE2016</i>	KU705239	coI	Roudbar, A.J.	Iran
<i>A. sp HRE2016</i>	KU705238	coI	Roudbar, A.J.	Iran
<i>A. sp HRE2016</i>	KU705237	coI	Roudbar, A.J.	Iran
<i>A. orhidanus</i>	KJ552755	coI	Geiger, M.F.	Albania
<i>A. orhidanus</i>	KJ552501	coI	Geiger, M.F.	Albania
<i>A. orhidanus</i>	KJ552730	coI	Geiger, M.F.	Albania
<i>A. cf prespensis</i>	KJ552526	coI	Geiger, M.F.	Greece
<i>A. cf prespensis</i>	HQ600665	coI	Triantafyllidis, A.,	Greece
<i>A. cf prespensis</i>	KJ552408	coI	Geiger, M.F.	Greece
<i>A. eichwaldii</i>	KU705240	coI	Roudbar, A.J.	Iran
<i>A. eichwaldii</i>	KU705241	coI	Roudbar, A.J.	Iran
<i>A. eichwaldii</i>	KU705242	coI	Roudbar, A.J.	Iran
<i>A. namaki</i>	KU705251	coI	Roudbar, A.J.	Iran
<i>A. namaki</i>	KU705252	coI	Roudbar, A.J.	Iran
<i>A. namaki</i>	KU705253	coI	Roudbar, A.J.	Iran
<i>A. coadi</i>	KU705256	coI	Roudbar, A.J.	Iran
<i>A. coadi</i>	KU705257	coI	Roudbar, A.J.	Iran
<i>A. coadi</i>	KU705258	coI	Roudbar, A.J.	Iran
<i>A. tabarestanensis</i>	KU705267	coI	Roudbar, A.J.	Iran
<i>A. tabarestanensis</i>	KU705268	coI	Roudbar, A.J.	Iran
<i>A. tabarestanensis</i>	KU705269	coI	Roudbar, A.J.	Iran
<i>A. samii</i>	KU705272	coI	Roudbar, A.J.	Iran
<i>A. samii</i>	KU705271	coI	Roudbar, A.J.	Iran
<i>A. idignensis</i>	KU705247	coI	Roudbar, A.J.	Iran
<i>A. idignensis</i>	KU705248	coI	Roudbar, A.J.	Iran
<i>A. idignensis</i>	KU705249	coI	Roudbar, A.J.	Iran
<i>A. nicolausi</i>	KU705259	coI	Roudbar, A.J.	Iran
<i>A. nicolausi</i>	KU705260	coI	Roudbar, A.J.	Iran
<i>A. nicolausi</i>	KU705261	coI	Roudbar, A.J.	Iran
<i>A. qanati</i>	KU705262	coI	Roudbar, A.J.	Iran
<i>A. qanati</i>	KU705266	coI	Roudbar, A.J.	Iran
<i>A. qanati</i>	KU705265	coI	Roudbar, A.J.	Iran
<i>A. holciki</i>	KU705244	coI	Roudbar, A.J.	Iran
<i>A. holciki</i>	KU705245	coI	Roudbar, A.J.	Iran
<i>A. holciki</i>	KU705246	coI	Roudbar, A.J.	Iran
<i>A. fangfangae</i>	KJ552616	coI	Geiger, M.F.	Albania
<i>A. fangfangae</i>	KJ552562	coI	Geiger, M.F.	Albania
<i>A. fangfangae</i>	KJ552720	coI	Geiger, M.F.	Albania
<i>A. devolli</i>	KJ552370	coI	Geiger, M.F.	Albania
<i>A. devolli</i>	KJ552420	coI	Geiger, M.F.	Albania
<i>A. devolli</i>	KJ552652	coI	Geiger, M.F.	Albania
<i>A. thessalicus</i>	KJ552656	coI	Geiger, M.F.	Greece

A. thessalicus	KJ552723	coI	Geiger,M.F.	Greece
A. thessalicus	KJ552369	coI	Geiger,M.F.	Greece
A. b.strymonicus	KJ552521	coI	Geiger,M.F.	Greece
A. b.strymonicus	KJ552519	coI	Geiger,M.F.	Greece
A. b.strymonicus	KJ552603	coI	Geiger,M.F.	Greece



### Appendix 3.3 Table of downloaded recombination activating gene I sequences:

Species	Accession number	gene	Authors	Country
<i>A. bipunctatus</i>	HM560384	RAGI	Perea, S.	France
<i>A. bipunctatus</i>	KM874706	RAGI	Stierandova,S.,	Czech Republic
<i>A. bipunctatus</i>	KM874703	RAGI	Stierandova,S.,	Poland
<i>A. bipunctatus rossicus</i>	KM874721	RAGI	Stierandova,S.,	Ukraine
<i>A. bipunctatus rossicus</i>	KM874719	RAGI	Stierandova,S.,	Russia
<i>A. tzanevi</i>	KM874726	RAGI	Stierandova,S.,	Bulgary
<i>A. tzanevi</i>	KM874725	RAGI	Stierandova,S.,	Bulgary
<i>A. tzanevi</i>	KM874724	RAGI	Stierandova,S.,	Bulgary
<i>A. sp 2</i>	KM874733	RAGI	Stierandova,S.,	Romania
<i>A. sp 2</i>	KM874723	RAGI	Stierandova,S.,	Bulgary
<i>A. sp 1</i>	KM874728	RAGI	Stierandova,S.,	Bosnia and Herzegovina
<i>A. sp 1</i>	KM874701	RAGI	Stierandova,S.,	Hungary
<i>A. orhidanus</i>	KM874707	RAGI	Stierandova,S.,	Albania
<i>A. cf prespensis</i>	KM874727	RAGI	Stierandova,S.,	Albania
<i>A. cf prespensis</i>	KM874700	RAGI	Stierandova,S.,	Albania
<i>A. cf prespensis</i>	KM874699	RAGI	Stierandova,S.,	Albania
<i>A. sp 3</i>	KM874737	RAGI	Stierandova,S.,	Greece
<i>A. sp 3</i>	KM874736	RAGI	Stierandova,S.,	Macedonia
<i>A. sp 3</i>	KM874734	RAGI	Stierandova,S.,	Greece
<i>A. fasciatus</i>	KM874717	RAGI	Stierandova,S.,	Russia
<i>A. fasciatus</i>	KM874716	RAGI	Stierandova,S.,	Russia
<i>A. fasciatus</i>	KM874715	RAGI	Stierandova,S.,	Russia
<i>A. kubanicus</i>	KM874708	RAGI	Stierandova,S.,	Russia
<i>A. sp 5</i>	KM874714	RAGI	Stierandova,S.,	Azerbaijan
<i>A. sp 5</i>	KM874711	RAGI	Stierandova,S.,	Azerbaijan
<i>A. sp 5</i>	KM874710	RAGI	Stierandova,S.,	Azerbaijan
<i>A. eichwaldii</i>	KM874712	RAGI	Stierandova,S.,	Azerbaijan
<i>A. eichwaldii</i>	KM874709	RAGI	Stierandova,S.,	Azerbaijan
<i>A. b.strymonicus</i>	KM874741	RAGI	Stierandova,S.,	Greece
<i>A. b.strymonicus</i>	KM874740	RAGI	Stierandova,S.,	Greece
<i>A. b.strymonicus</i>	KM874739	RAGI	Stierandova,S.,	Bulgaria